

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ANALYSE STATISTIQUE DE PLANS PAR
REGROUPEMENT ET APPLICATIONS À LA GÉNOMIQUE
ET À LA BIOSTATISTIQUE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

PIERRE AUGER

JANVIER 2010

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 -Rév.01-2006). Cette autorisation stipule que «conformément à l'article **11** du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

TABLE DES MATIÈRES

LISTE DES FIGURES	v
LISTE DES TABLEAUX	vii
LISTE DES SYMBOLES	ix
RÉSUMÉ	xi
INTRODUCTION	1
CHAPITRE I	
PRÉALABLE	3
1.1 Introduction	3
1.2 Éléments de génétique	3
1.3 Criblage de banques génomiques	12
CHAPITRE II	
MÉTHODE SIMPLICIALE	21
2.1 Introduction	21
2.2 Préliminaires	22
2.2.1 Graphes et complexes simpliciaux	24
2.2.2 Monoïdes de matrices booléennes	27
2.3 Plans d -disjoints	30
2.4 Matrices d'incidence associées à un complexe simplicial	37
CHAPITRE III	
PLANS ORDONNÉS	43
3.1 Introduction	43
3.2 Plans ordonnés	43
3.3 Détermination de la probabilité à l'aide de la méthode ensembliste	51
3.4 Lois de probabilités associées à l'espèce des plans par regroupement	58
CHAPITRE IV	
MISE EN APPLICATION DE L'ESTIMATEUR DE MONTE CARLO DE L'ES- PÉRANCE DE LA STATISTIQUE Σ	59

4.1	Introduction	59
4.2	Notions préliminaires	59
4.3	Criblage de banque à l'aide de la méthode de reconfiguration	63
4.3.1	Condition d'acceptabilité	65
4.3.2	Exemple de reconfiguration	75
4.4	Les grandes étapes	77
4.5	Approximations par arrondi à droite de la loi binomiale	80
4.6	Maximisation du domaine d'évaluabilité	85
CHAPITRE V		
	SIMULATIONS	91
5.1	Introduction	91
5.2	Tables	91
CONCLUSION		97
BIBLIOGRAPHIE		99

LISTE DES FIGURES

1.1	Matrice d'incidence de \mathcal{P} par rapport au couple (φ, ψ)	18
1.2	Tableau des valeurs des échantillons : microplaque de 96 cupules	20
2.1	Une représentation planaire de g	23
2.2	Le graphe complet à 6 sommets	23
2.3	Un couplage du graphe complet à 6 sommets	24
2.4	Un ensemble saturant de sommets	41
4.1	Un réseau carré avec $l = 3$	62
4.2	Configuration hypercubique de paramètres 2, 3	65
4.3	Nouvelle configuration	65
4.4	Configuration initiale	76
4.5	Deuxième configuration	76
4.6	Troisième configuration	77

LISTE DES TABLEAUX

2.1	Parties de cardinal 2 de $[4]$	36
2.2	Indices correspondant aux couples $(j, I) \in \bigcup_{j \in [4]} \{j\} \times \mathcal{P}_2([4] \setminus \{j\})$. . .	36
4.2	Mineurs de M	70
5.2	$d = 2$	93
5.2	$d = 3$	94
5.2	$d = 4$	94
5.2	$d = 5$	95
5.2	$d = 6$	95

LISTE DES SYMBOLES

Tout au long du présent ouvrage, nous noterons $[n]$, pour n un nombre naturel arbitraire, l'ensemble formé des nombres naturels allant de 1 à n , c'est-à-dire $[n] = \{1, 2, \dots, n\}$. De façon analogue, nous noterons $\llbracket n \rrbracket$ l'ensemble $\{0, 1, \dots, n-1\}$.

Soit E un ensemble fini. Nous noterons $\mathcal{P}(E)$ l'ensemble des parties de E . Plus particulièrement, pour k un nombre naturel, nous noterons $\mathcal{P}_k(E)$ l'ensemble des parties à k éléments de E . Par abus de notation, nous noterons, afin d'alléger les notations, $\mathcal{P}(n)$ l'ensemble des parties de $[n]$, pour n un nombre naturel arbitraire, alors qu'on devrait plutôt écrire, si on se conforme à la notation précédente, $\mathcal{P}([n])$.

\mathbf{N}_+ : ensemble des nombres naturels strictement positifs

\mathbb{B} : catégorie des ensembles finis et bijections

$\mathbf{C}_n(\mathfrak{B})$: monoïde des vecteurs colonnes à n dimensions sur \mathfrak{B}

\mathfrak{P} : ensemble des nombres premiers

\mathbb{F} : classe des ensembles finis

RÉSUMÉ

Dans ce mémoire, nous présentons quelques facettes de l'étude statistique des plans par regroupement, qui nous ont paru particulièrement représentatives. Comme on rencontre les méthodologies faisant appel aux plans par regroupement surtout en génétique, nommément en génétique expérimentale, où du reste ces méthodologies s'y révèlent remarquablement adéquates, nous présentons d'abord en le chapitre I, en guise de préalable à la suite du mémoire, un certain nombre d'éléments de génétique et, en particulier, nous y présentons la notion de sonde, qui est en quelque sorte une « constante », puisque cette notion est sous-jacente à toutes les méthodologies que nous présentons ici, au sens où, plus précisément, cette notion est centrale dans ce qui motive toutes ces méthodologies.

Dans le chapitre 2, nous présentons le concept de complexe simplicial. Ce concept a ceci de remarquable qu'il donne lieu à des plans par regroupement grâce auxquels on peut déterminer tous les éléments positifs à partir de la donnée des groupes positifs.

Nous introduisons ensuite dans le chapitre 3 une nouvelle espèce de structures, à savoir l'espèce des plans ordonnés. Nous en déduisons ensuite une formule pour l'espérance du nombre d'éléments négatifs indéterminés dans le cas des plans par regroupement ordinaires. Nous présentons aussi une méthode, que nous qualifions d'ensembliste, pour la détermination de probabilités toujours dans le cas des plans ordonnés. Nous pensons avoir détecté une faille dans le cours du raisonnement de l'auteur de cette méthode et en proposons une solution de rechange.

Dans le chapitre 4, nous présentons une méthode d'estimation de l'espérance d'un certain estimateur du nombre d'éléments positifs, estimateur que nous y définissons, pour une classe particulière de plans par regroupement, à savoir la classe des plans hypercubiques. Nous présentons ensuite différentes facettes de la mise en application de l'estimateur de Monte Carlo de l'espérance de cet estimateur-là. Cette mise en application nous permet finalement dans le chapitre 5 d'obtenir des tables d'estimations de cet estimateur pour différentes valeurs des paramètres de cette classe de plans, et pour différentes valeurs du paramètre de la loi de probabilité associée.

Mots-clé : génomique, complexe simplicial, plan par regroupement, plan ordonné, plan hypercubique.

INTRODUCTION

L'idée consistant à vouloir regrouper des éléments d'un ensemble dont on souhaite extraire la partie formée des éléments vérifiant une certaine relation – nous dirons dans ce qui suit que ce sont là les éléments « défectueux » – apparaît naturellement dans certaines situations où le nombre de ces éléments défectueux est petit par rapport au nombre total d'éléments, d'une part, et où, d'autre part, le coût associé à l'analyse d'un certain nombre d'éléments arbitraires pris en bloc non seulement n'est pas proportionnel au nombre de ces éléments, mais est en fait inférieur au coût total associé à l'analyse de chacun de ces éléments pris un à un.

Plus précisément, supposons qu'on dispose d'un moyen permettant uniquement de déterminer si une partie arbitraire d'un ensemble d'objets contient ou non un ou plusieurs éléments défectueux, c'est-à-dire que, dans le cas où il y a détection d'un ou plusieurs éléments défectueux, ce même moyen ne dit pas lesquels le sont. On fait alors face au dilemme suivant : si on trouve qu'une partie ne contient aucun élément défectueux, alors on réalise une économie qui, dans certains cas, peut être majeure, alors que, dans le cas contraire, on ne dispose pas de toute l'information dont on disposerait en faisant une analyse élément par élément.

Historiquement, un des premiers exemples de solution mathématique à un problème de ce type (Dorfman, 1943) est apparu dans le contexte de la deuxième guerre mondiale, en les États-Unis. L'auteur était motivé par le dilemme suivant. Il s'agissait de déterminer si chacune des recrues était porteuse ou non de la syphilis. En mélangeant ensemble les échantillons de sérum provenant de plusieurs recrues, on obtient évidemment un nouvel échantillon indivisible, que nous appellerons échantillon composé. Si aucune des recrues dont provient un tel échantillon n'est porteuse de la maladie, alors une analyse révélera l'absence de la maladie pour toutes ces mêmes recrues, d'où une économie,

puisque le coût associé à l'analyse d'un échantillon est le même que cet échantillon soit simple ou composé. Dans le cas contraire, c'est-à-dire si au moins une des recrues dont provient l'échantillon composé est porteuse de la maladie, on ne peut déterminer laquelle sans recourir à d'autres analyses. Dans ce cas-ci, on ne réalise donc pas l'économie escomptée.

Supposons par exemple qu'une recrue sur 100 soit porteuse de la maladie et que l'analyse d'un échantillon coûte 1 unité monétaire, que l'échantillon soit simple ou non. Supposons de plus que la population de recrues soit de 1000 individus. Il en coûterait donc 1000 unités monétaires pour l'analyse des 1000 échantillons simples. Voyons maintenant combien il en coûterait dans l'hypothèse d'un plan par regroupement. Considérons par exemple le plan consistant à regrouper les recrues par groupes de 100 individus. Nous allons supposer que l'espérance du nombre de groupes contenant au moins un individu porteur est 5. Par suite, nous allons supposer qu'en fait ce nombre de groupes est 5, pour une réalisation du plan que nous allons considérer maintenant. Ainsi, on trouve que 500 recrues sont saines. Le coût associé s'élève donc à 5 unités monétaires au lieu de 500 unités monétaires, ce qui est plus qu'appréciable. On a donc éliminé 500 recrues du processus. Même en analysant les 500 recrues restantes une à la fois, on aboutit à une économie considérable, puisque le coût total s'élèverait dans ce cas à $500 + 5 = 505$ analyses, au lieu de 1000. En fait, une stratégie qui nous semble préférable consisterait à itérer ce procédé à la population restante, en élaborant éventuellement un nouveau plan formé de groupes plus petits, par exemple en formant des groupes de 50 au lieu de 100. On se convainc sans trop de peine qu'un tel processus aboutit en quelques étapes seulement, et que le coût total est, dans tous les cas, nettement inférieur à 500 unités monétaires.

CHAPITRE I

PRÉALABLE

1.1 Introduction

Le type d'opérations visant à déterminer la présence ou l'absence d'une sonde donnée dans chacun des segments d'ADN pris un à un d'une banque génomique est un des types d'expérimentations les plus courants en génétique expérimentale. Pour en savoir plus sur la nature de ces opérations en génétique expérimentale, on pourra consulter par exemple l'ouvrage de Bernot (Bernot, 1996) où on trouve l'essentiel sur la question, ou encore, pour un exposé plus détaillé, l'ouvrage de Brown (Brown, 2004). Pour ce qui est du point de vue abstrait, nous recommandons l'ouvrage d'Autebert (Autebert, 1987), pour ce qui est de la théorie des langages, et l'ouvrage de Bourbaki (Bourbaki, 1970), pour ce qui est des notions plus proprement algébriques.

1.2 Éléments de génétique

Rappelons d'abord qu'un alphabet n'est rien d'autre qu'un ensemble fini de symboles. Tout au long de la présente section, \mathcal{A} désignera un alphabet quelconque.

Définition 1. Soit \mathcal{A} un alphabet, et soit $n \in \mathbf{N}$ quelconque. Un mot sur \mathcal{A} est une application $m : [n] \rightarrow \mathcal{A}$. On dit alors de n que c'est la longueur de m . De plus, on note \mathcal{A}^n l'ensemble des mots de longueur n sur \mathcal{A} . Finalement, pour $i \in [n]$, on pose $m_i = m(i)$.

Le plus souvent, on écrit un mot donné (au sens technique de ce terme, que nous venons tout juste de donner dans la définition ci-dessus) en juxtaposant les images respectives des nombres naturels du domaine de ce mot, dans l'ordre croissant des éléments du domaine, de gauche à droite. Par exemple, posons $\mathcal{A} = \{A, C, G, T\}$ et soit $m : [4] \rightarrow \mathcal{A}$ le mot défini par $m_1 = C, m_2 = C, m_3 = A, m_4 = T$. On écrit alors $m = \text{CCAT}$.

Définition 2. Soit \mathcal{A} un alphabet. On note alors \mathcal{A}^* l'ensemble de tous les mots sur \mathcal{A} . Plus précisément, on définit \mathcal{A}^* en posant

$$\mathcal{A}^* = \bigsqcup_{n \in \mathbb{N}} \mathcal{A}^n.$$

On dit alors d'un élément quelconque de \mathcal{A}^* que c'est un mot sur \mathcal{A} .

Définition 3. Une séquence d'ADN est un mot sur $\mathcal{A} = \{A, C, G, T\}$.

Définition 4. Soit \mathcal{A} un alphabet. Soit f la loi de composition sur \mathcal{A}^* qu'on définit, pour $u, v \in \mathcal{A}^*$, par

$$(f(u, v))_i = \begin{cases} u_i, & \text{si } 0 \leq i \leq |u|, \\ v_{i-|u|} & \text{si } |u| < i \leq |u| + |v|. \end{cases}$$

On dit alors de cette loi que c'est la loi de concaténation sur \mathcal{A}^* .

Autrement dit, le composé de deux mots u et v (dans cet ordre) est le mot qu'on obtient en juxtaposant u et v , u étant à gauche et v à droite. On dit alors de ce composé qu'on l'obtient en concaténant u et v .

Exemple 1. Posons $\mathcal{A} = \{A, C, G, T\}$. Soit $u = \text{CCGAT}$, $v = \text{ACGATG}$. Alors, le composé de u et v (dans cet ordre), noté uv , est le mot CCGATACGATG .

Proposition 1.0. — L'ensemble \mathcal{A}^* muni de la loi de concaténation est un monoïde.

Démonstration. En effet, on vérifie aisément que cette loi de composition est associative, et que le mot vide en est l'élément neutre. \square

Le monoïde mentionné dans la proposition ci-dessus est appelé monoïde libre engendré par \mathcal{A} . Désormais, par abus de langage, nous désignerons ce monoïde par \mathcal{A}^* .

Dans ce qui suit, nous allons noter $\mathcal{J}(\mathbf{N})$ l'ensemble des parties de \mathbf{N} de la forme $\{n \in \mathbf{N} \mid a < n < b\}$, où $a, b \in \mathbf{N}$, $a \leq b$. Cet ensemble n'est autre que l'analogue discret de l'ensemble des intervalles dans \mathbf{R} . Plus particulièrement, on se convainc aisément que tout $I \in \mathcal{J}(\mathbf{N})$ est l'intersection d'un intervalle dans \mathbf{R} et \mathbf{N} , c'est-à-dire est un ensemble de la forme $J \cap \mathbf{N}$, où J est un intervalle dans \mathbf{R} . Par exemple, posons $I = \{2, 3\}$. On a alors $I = [2, 3] \cap \mathbf{N}$, où $[2, 3]$ désigne l'intervalle fermé d'extrémités 2 et 3. Notons $\mathcal{C}_2(\mathbf{N})$ l'ensemble des couples $(a, b) \in \mathbf{N} \times \mathbf{N}$ tels que $a \leq b$. Rappelons qu'on a une bijection

$$\begin{aligned} \beta: \mathcal{J}(\mathbf{N}) &\xrightarrow{\sim} \mathcal{C}_2(\mathbf{N}) \\ I &\mapsto (\min(I), \max(I)). \end{aligned}$$

Soit $I \in \mathcal{J}(\mathbf{N})$. Posons $a = \min(I)$, $b = \max(I)$. On note alors également $I_{a,b}$ cet intervalle.

Définition 5. Soit $\delta: \mathcal{J}(\mathbf{N}) \times \mathbf{N} \rightarrow \mathcal{J}(\mathbf{N})$ la loi de composition externe sur $\mathcal{J}(\mathbf{N})$ ayant \mathbf{N} comme ensemble d'opérateurs, définie, pour $(I, n) \in \mathcal{J}(\mathbf{N}) \times \mathbf{N}$ quelconque, par $\delta(I_{a,b}, n) = I_{a+n, b+n}$. On dit alors de l'opérateur f_n associé à $n \in \mathbf{N}$ que c'est l'opérateur de décalage à droite de n positions. On écrit alors, pour $(I, n) \in \mathcal{J}(\mathbf{N}) \times \mathbf{N}$ quelconque, $f_n(I) = I + n$.

Soit $I \in \mathcal{J}(\mathbf{N})$, $n \in \mathbf{N}$. La notation $I + n$ suggère alors simplement le fait qu'on décale « uniformément » l'intervalle I de n positions vers la droite.

Exemple 2. Posons $I = [2]$, $n = 3$. On a donc, par définition, $f_3(I) = I + 3$, c'est-à-dire $f_3(I) = I_{4,5} = \{4, 5\}$.

Définition 6. Soit $I \in \mathcal{J}(\mathbf{N})$ quelconque. Posons $k = |I|$. On note alors ψ_I l'unique bijection croissante de $[k]$ sur I .

On a donc là une famille $(\psi_I)_{I \in \mathcal{J}(\mathbf{N})}$ de bijections.

Définition 7. Soit $n \in \mathbf{N}$. On note alors $\mathcal{J}(n)$ l'ensemble défini par

$$\mathcal{J}(n) = \{I \in \mathcal{J}(\mathbf{N}) \mid I \subseteq [n]\}.$$

Définition 8. Soit $m \in \mathcal{A}^*$ quelconque. Posons $n = |m|$. Soit $I \in \mathcal{J}(n)$ quelconque. On dit alors de la restriction de m à I , notée m_I , que c'est un segment de m . De plus, posons $k = |I|$, $i = \min(I)$. On dit alors que k est la longueur de m_I et que i en est la position initiale.

Définition 9. Soit $m \in \mathcal{A}^*$ quelconque. Posons $n = |m|$. Soit $m' \in \mathcal{A}^*$ quelconque. Soit $I \in \mathcal{J}(n)$. On dit alors que m_I est une occurrence de m' dans m si $m' = m_I \circ \psi_I$.

Exemple 3. Posons $m = \text{ATGAAAAGGGACCACAGGCTCCCTCAC}$. Alors, on vérifie aisément que le segment de m de position initiale 16 et de longueur 7 est une occurrence du mot AGGCTCC dans m .

Définition 10. Soit $m \in \mathcal{A}^*$, et soit \mathcal{S}_m l'ensemble des segments de m . Soient $s, s' \in \mathcal{S}_m$ quelconques. Soient $I, I' \in \mathcal{J}(\mathbb{N})$ tels que $s = m_I$, $s' = m_{I'}$. On définit alors la relation d'équivalence sur \mathcal{S}_m notée \sim en posant $s \sim s'$ si, et seulement si, $s \circ \psi_I = s' \circ \psi_{I'}$.

Autrement dit, deux segments sont équivalents pour la relation que nous venons de définir si, et seulement si, ces segments sont des occurrences d'un même mot. On voit donc que chacune des classes d'équivalence pour cette relation est formée de toutes les occurrences d'un même mot.

Définition 11. On note $\nu_m(s)$ le nombre d'occurrences du mot associé à s dans m .

Exemple 4. Le mot suivant, qui est une séquence d'ADN, c'est-à-dire une séquence de nucléotides, comporte deux occurrences du mot CGGGCAGGAGGGGGAGG :

```

... ..TCAGGGTGAGAAGGATGAAAAGGGACCACAGGCTCCCTCACCCCTTACC
GTGGGCCAAATGCTTGACCTGGGTGGCAGTGAGTGGGCGGGTAATCGGGCAG
GAGGGGGAGGCGGGCAGGAGGGGGGAGGCGGGCACGAGGGGGGAGGTGAGC
AGGAGGGGGAGGCGGGCAGGAGGAGGCGGGCAGGAGGGGGAGACGGGC
AGGAGAGGGAGGCAGGCAGGAGAGGGAGGTGGGCAGGAGGGGGGGCGGGC
AGGAGGAGGAGGTGGGCAGGAGGGGGAGGCGGGCAGGAGGGGGAGGCGGGC
AGGAGGGTGAGGGGGGATCTGGACGCCCGGGGAGACTGAGGGAGGATCCAAG

```

CCCCAGGGCTCCTTGAGGAAACAACAGGGGTGCCAGACGTGGCCCGGGCCCCCT
 GGCTGGGCCAGTTCGGGGTGTGTGGGAGCTGAGGACTCACTGGGCTTGAGGA
 CTGACTGATGTGGA:.....

Définition 12. Soit M un ensemble de mots sur \mathcal{A} , c'est-à-dire $M \subseteq \mathcal{P}(\mathcal{A}^*)$. Soit $m \in \mathcal{A}^*$ quelconque. On dit que m est un site STS de M si $\sum_{m \in M} \nu_m(s) = 1$.

Autrement dit, un segment est un site STS d'un ensemble de mots M s'il existe un mot $m_0 \in M$ tel qu'on ait

$$(\nu_{m_0}(s) = 1) \wedge (\forall m \in M \setminus \{m_0\})(\nu_m(s) = 0).$$

L'intérêt d'un site STS est qu'il permet de repérer un locus de façon univoque (Olson *et al.*, 1989).

Soit H un sous-groupe du groupe $\mathfrak{S}(\mathcal{A})$ des permutations de \mathcal{A} . Soit $\alpha_H : H \times \mathcal{A}^* \rightarrow \mathcal{A}^*$ l'action de groupe associée. Rappelons qu'on note \mathcal{A}^*/H l'ensemble quotient de \mathcal{A}^* par la relation de conjugaison associée à α_H .

Soit Π un ensemble de permutations de \mathcal{A} , c'est-à-dire qu'on a $\Pi \subseteq \mathfrak{S}(\mathcal{A})$. Rappelons qu'on note alors $\langle \Pi \rangle$ le groupe engendré par Π . Plus particulièrement, si $|\Pi| = 1$, on note alors, par abus de langage, $\langle \sigma \rangle$, le groupe engendré par Π , lorsqu'on pose $\Pi = \{\sigma\}$.

Définition 13. Soit $\mathcal{A} = \{A, C, G, T\}$. Soit $\sigma \in \mathfrak{S}(\mathcal{A})$ définie par $\sigma = (AT)(CG)$, c'est-à-dire que σ est la permutation de \mathcal{A} dont la décomposition cyclique est formée des deux transpositions (AT) et (CG). On dit alors d'un élément quelconque de l'ensemble quotient $\mathcal{A}^*/\langle \sigma \rangle$ que c'est une molécule d'ADN.

Cette définition exprime entre autres choses le fait qu'une molécule d'ADN est « double brin »

Exemple 5. L'ensemble formé des séquences d'ADN

$$\text{ATGAAAAGGGACCACAGGCTCCCTCAC} \quad (1.1)$$

et

$$\text{TACTTTTCCCTGGTGTCCGAGGGAGTG} \quad (1.2)$$

est une molécule d'ADN.

Définition 14. Soit $m \in \mathcal{A}^*/\langle \sigma \rangle$ quelconque. On note alors $\mathcal{S}(m)$ l'ensemble défini par

$$\mathcal{S}(m) = \bigcup_{\gamma \in m} \mathcal{S}(\gamma).$$

On dit alors d'un élément quelconque de $\mathcal{S}(m)$ que c'est un segment de m .

On voit donc que l'ensemble des segments d'une molécule d'ADN s'obtient en réunissant les ensembles de segments respectifs des deux brins dont elle est formée.

Exemple 6. Revenons à l'exemple précédent. Soit m la molécule d'ADN formée des deux chaînes (1.1) et (1.2). On obtient alors l'ensemble des segments de m en réunissant l'ensemble des segments de la chaîne (1.1) et l'ensemble des segments de la chaîne (1.2).

Proposition 1.1. – Soit $\mathcal{O} \in \mathcal{A}^*/\langle \sigma \rangle$. On a $|\mathcal{O}| = 2$.

Démonstration. En effet, on vérifie aisément que l'application α_σ est une involution ne laissant fixe aucun élément, c'est-à-dire que, pour tout $m \in \mathcal{A}^*$, on a $\alpha_\sigma(m) \neq m$. \square

En particulier, cette proposition exprime le fait que les deux brins d'une molécule d'ADN quelconque sont distincts en tant que mots sur \mathcal{A} .

Considérons une fonction de la forme $f : \mathcal{A}^* \rightarrow \mathcal{A}^*$ quelconque. En particulier, pour $m \in \mathcal{A}^*$ quelconque, l'image par f de m est un mot sur \mathcal{A} . Dans ce qui suit, nous noterons $(f(m))_i$, pour $i \in [|m|]$, la lettre occupant la i -ième position de $f(m)$.

Définition 15. Soit $\sigma \in \mathfrak{S}(\mathcal{A})$. On note alors f_σ l'application de \mathcal{A}^* dans \mathcal{A}^* telle que, pour $m \in \mathcal{A}^*$ quelconque, on a, quel que soit $i \in [|m|]$,

$$(f_\sigma(m))_i = \sigma(m_i), \quad i = 1, 2, \dots, |m|.$$

Exemple 7. Soit $m = \text{ATGAAAAGGGACCACAGGCTCCCTCAC}$. Posons $\sigma = (\text{AT})(\text{CG})$.

On a alors

$$f_\sigma(m) = \text{TACTTTTCCCTGGTGTCCGAGGGAGTG}.$$

Définition 16. Soit $m \in \mathcal{A}^*$, et soit $s \in \mathcal{S}(m)$ quelconque. On définit alors comme suit l'application $\mu_m : \mathcal{S}(m) \rightarrow \mathcal{A}^*$. Soit I le domaine de s . Le mot $\mu_m(s)$ est alors défini par $\mu_m(s) = m \circ \psi_I$.

Exemple 8. Posons $m = \text{ATGAAAAGGGACCACAGGCTCCCTCAC}$. Soit s le segment de m de domaine $[3] + 10$. On a alors $\mu_m(s) = \text{ACC}$.

Définition 17. Soit \mathcal{G} un ensemble de molécules d'ADN. On dit alors que \mathcal{G} est un génome.

Dans ce qui suit, nous noterons \mathcal{G} un génome quelconque.

Définition 18. Soit \mathcal{G} un génome. On note alors $\mathcal{S}(\mathcal{G})$ l'ensemble défini par

$$\mathcal{S}(\mathcal{G}) = \bigcup_{m \in \mathcal{G}} \mathcal{S}(m). \quad (1.3)$$

On dit alors d'un élément quelconque de cet ensemble que c'est un segment de \mathcal{G} .

Définition 19. Soit \mathcal{G} un génome. On note alors $\mathcal{S}'(\mathcal{G})$ l'ensemble défini par

$$\mathcal{S}'(\mathcal{G}) = \bigsqcup_{m \in \mathcal{G}} \bigcup_{\gamma \in m} \mathcal{S}(\gamma). \quad (1.4)$$

Définition 20. Soit \mathcal{G} un génome. On dit alors d'un élément quelconque de $\mathcal{S}'(\mathcal{G})$ que c'est un segment génomique numéroté.

Il peut très bien arriver que deux chaînes distinctes γ et γ' associées respectivement à deux molécules distinctes admettent un même segment, c'est-à-dire soient telles qu'on ait $\gamma \circ \psi_I = \gamma' \circ \psi_I$, pour un certain $I \in \mathcal{I}$. En ce cas, après réunion des ensembles $\mathcal{S}(\gamma)$ et $\mathcal{S}(\gamma')$, ces deux segments sont indiscernables, c'est-à-dire ne donnent lieu qu'à

un élément. En revanche, deux tels segments conservent leur individualité après réunion disjointe des mêmes deux ensembles, c'est-à-dire donnent lieu à deux éléments distincts de cette réunion disjointe.

Proposition 1.2. – Soit \mathcal{G} un génome. Soit $\beta : [|\mathcal{G}|] \xrightarrow{\sim} \mathcal{G}$ quelconque. On a

$$|C_\beta(\mathcal{G})| = 2 \sum_{m \in \mathcal{G}} |\mathcal{S}(m)|.$$

Le cardinal de l'ensemble $C_\beta(\mathcal{G})$ étant indépendant de la bijection β considérée, nous noterons $c(\mathcal{G})$ ce cardinal.

Soit β une numérotation de \mathcal{G} . Dans ce qui suit, nous noterons $\mathcal{E}_\beta(\mathcal{G})$ l'ensemble probabilisé qu'on obtient en munissant $\mathcal{S}_\beta(\mathcal{G})$ de la loi de probabilité uniforme.

Définition 21. Soit \mathcal{G} un génome. Soient $m \in \mathcal{G}$, $\gamma \in m$. Soit $n \in \mathbf{N}$. On note alors $\mathcal{S}_n(\gamma)$ l'ensemble des segments de γ de longueur n .

Proposition 1.3. – On a

$$|\mathcal{S}_n(\gamma)| = \begin{cases} |\gamma| - n + 1, & \text{si } n \leq |\gamma|, \\ 0, & \text{si } n > |\gamma|. \end{cases}$$

On pose alors $s_n(\gamma) = |\mathcal{S}_n(\gamma)|$.

Démonstration. Facile. □

Définition 22. Soit \mathcal{G} un génome. Soit $n \in \mathbf{N}$. On note alors, par abus de langage, $C_n(\mathcal{G})$ l'ensemble des β -segments génomiques de longueur n , la bijection β étant arbitraire. On pose alors $c_n(\mathcal{G}) = |C_n(\mathcal{G})|$.

Proposition 1.4. – Soit une famille $(\gamma_m)_{m \in \mathcal{G}}$ telle que, pour $m \in \mathcal{G}$, $\gamma_m \in m$. On a alors

$$c_n(\mathcal{G}) = 2 \sum_{m \in \mathcal{G}} s_n(\gamma_m).$$

Démonstration. Cette identité résulte de ce que, quel que soit $m \in \mathcal{G}$, on a

$$\bigcap_{\gamma \in m} \mathcal{S}(\gamma) = \emptyset,$$

de sorte que $\mathcal{S}(m) = \mathcal{S}(\gamma) \sqcup \mathcal{S}(\gamma')$, lorsqu'on pose $m = \{\gamma, \gamma'\}$. \square

Proposition 1.5. – Soit s un site STS de \mathcal{G} . Posons $l = |s|$.

Définition 23. Soient $m, m' \in \mathcal{A}^*$ quelconques. On note alors \preceq la relation binaire sur \mathcal{A}^* définie par

$$m \preceq m' \iff (\exists I \in \mathcal{I}(\mathbf{N})) (m' \circ \psi_I = m).$$

Exemple 9. Posons $m = \text{ATTCGAA}$, $m' = \text{ACAACAACATTCGAA}$. On a alors $m \preceq m'$. En effet, posons $I = [7] + 10$. On vérifie alors que $m' \circ \psi_I = m$.

Proposition 1.6. – La relation \sim sur \mathcal{A}^* est réflexive, antisymétrique et transitive.

Démonstration. Facile. \square

Proposition 1.7. – Soit \mathcal{G} un génome. Soient $s, s' \in \mathcal{S}(\mathcal{G})$. On note alors \sim la relation binaire sur $\mathcal{S}(\mathcal{G})$ définie par

$$s \preceq s' \iff \mu(s) \preceq \mu(s').$$

Définition 24. Soit $s \in \mathcal{S}(\mathcal{G})$. Soit $s' \in \mathcal{S}(\mathcal{G})$. On définit alors une relation unaire, notée $R_s(\mathcal{G})$, par

$$s' \in R_s(\mathcal{G}) \iff s \preceq s'.$$

En génétique expérimentale, on appelle sonde le segment s auquel on attribue le rôle donné ci-dessus. Cela dit, du point de vue axiomatique, ce segment n'est rien de plus qu'un segment, c'est-à-dire que cette relation est définie quel que soit le segment qu'on considère. Du point de vue abstrait, l'emploi de ce mot est donc superflu.

Notons que, en génétique expérimentale, on rencontre très souvent ce type de relations dans le cas où le segment noté s ci-dessus est un STS.

Proposition 1.8. – Soit s un site STS de \mathcal{G} . Posons $k = |s|$. Notons i_0 la position initiale de s . Soit $\gamma \in \mathcal{G}$ l'unique molécule de \mathcal{G} telle qu'il existe $m \in \gamma$ telle que $s \preceq m$. Soit $n \in \mathbb{N}$. On a alors

$$|R_s(\mathcal{G}) \cap \mathcal{S}_n(\mathcal{G})| = \begin{cases} n - k + 1, & \text{si } (i_0 > n - k) \wedge (\gamma \geq i_0 + n - 1), \\ i_0, & \text{si } (i_0 \leq n - k) \wedge (\gamma \geq i_0 + n - 1), \\ |\gamma| - (i_0 + k - 1), & \text{si } (i_0 > n - k) \wedge (\gamma < i_0 + n - 1), \\ |\gamma| - (n - 1), & \text{si } (i_0 \leq n - k) \wedge (\gamma < i_0 + n - 1). \end{cases}$$

Corollaire. Soit s un site STS. Soit $n \in \mathbb{N}$. Notons alors $E_n(s)$ l'événement de $\mathcal{E}(\mathcal{G})$ défini par

$$E_n(s) = \{s' \in \mathcal{S}_n(\mathcal{G}) \mid s \preceq s'\}.$$

On a alors

$$P(E_n(s)) = \frac{|R_s(\mathcal{G}) \cap \mathcal{S}_n(\mathcal{G})|}{c_n(\mathcal{G})}.$$

1.3 Criblage de banques génomiques

Du point de vue formel, une banque génomique est un ensemble de séquences d'ADN. Le plus souvent, on appelle banque génomique un ensemble de séquences qu'on obtient par découpage des chaînes d'ADN d'un génome donné. Autrement dit, en connaissant l'ordre de découpage, on retrouverait le génome par concaténation des séquences. Bien entendu, cela est purement théorique, car, du point de vue expérimental, il serait extrêmement onéreux, si tant est qu'il soit possible, voire concevable, de conserver l'ordre des séquences. Ce que nous sommes simplement en train de dire, c'est que l'ensemble des séquences non seulement recouvre tout le génome, mais, plus précisément, le recouvre sans chevauchements. On a donc une factorisation du génome en facteurs dont on n'a

pas conservé l'ordre. On se gardera donc de croire que la seule donnée d'une banque génomique équivale au génome dont elle dérive.

Par exemple, considérons la chaîne d'ADN

GGAGGCGGGCACGAGGGGGGAGGTGAGCAGGAGGGGGAGGCGGG
CAGGAGGAGGAGGCGGGCAGGAGGGGGAGACGGGCAGGAGAGGG
AGGCAGGCAGGAGAGGGAGGTGGGCAGGAGGGGGGGGCGGGCAG
GAGGAGGAGGTGGGCA

Selon la définition informelle que nous avons donné ci-dessus, l'ensemble formé des séquences GGAGGCGGGCACG, AGGGGGGAGGTGAGCAGGAGGGGGAGG CGGGCAGGAGGAGGAGGCGGG, CAGGAGGGGGAGACGGGCAGGAGAGGGA GGCAGGC, AGGAGAGGGA, GGTGGGCAGGAGGGGGGGGCGGGCAGGAGGA GGAG, GTGGGCA est une banque génomique. En effet, en concaténant ces séquences dans l'ordre où ils sont énumérés ici, on retrouve la chaîne donnée, de sorte que, plus particulièrement, ces séquences ne se chevauchent pas. Cela dit, peut-être est-il bon de noter que, la banque génomique étant donnée, on ne sait pas que la première des séquences précédentes est le premier facteur de la chaîne d'ADN donnée, ni que la deuxième des séquences précédentes est le deuxième facteur de la chaîne d'ADN donnée, et ainsi de suite. Comme dans le cas présents il y a $6!$ ordres de factorisations, on ne peut pas retrouver la chaîne donnée à partir de la banque génomique. Plus généralement, si on se donne un ensemble de mots M de cardinal n , alors le fait de savoir qu'il existe une bijection $\varphi : M \xrightarrow{\sim} [n]$ telle qu'on ait $\gamma = \varphi^{-1}(1) \cdots \varphi^{-1}(n)$ ne permet pas de déterminer celle-ci.

Définition 25. Soit U un ensemble fini arbitraire. On définit alors l'espèce \mathcal{P} des parties d'un ensemble en posant $\mathcal{P}[U] = \mathcal{P}(U)$.

Définition 26. On appelle l'espèce $\mathcal{P} \circ \mathcal{P}$ espèce des plans par regroupement. On note alors \mathcal{R} cette espèce. Soit U un ensemble fini quelconque. On dit alors d'un élément quelconque de l'ensemble $\mathcal{R}[U]$ que c'est un plan par regroupement sur U . Soit alors $\mathcal{P} \in \mathcal{R}[U]$ quelconque. On dit alors d'un élément quelconque de \mathcal{P} que c'est un groupe.

Soit U un ensemble fini. On voit donc par là qu'un élément de l'ensemble $\mathcal{R}[U]$ n'est rien d'autre qu'un ensemble de parties de U .

Exemple 10. Posons $U = [3]$. Soit $\mathcal{P} = \{\{2\}, \{1, 2\}, \{1, 2, 3\}\}$. On a alors que \mathcal{P} est un plan par regroupement sur $[3]$, ce qu'on note $\mathcal{P} \in \mathcal{R}[3]$.

Définition 27. On appelle ensemble factoriel pour \mathcal{G} un ensemble $F \in \mathcal{A}^*$ vérifiant la relation suivante : il existe une partition Π de F et une bijection $\beta : \Pi \xrightarrow{\sim} |\Pi|$ et une famille de bijections $(\beta_i)_{i \in |\Pi|}$ tels que, pour $i \in [|\Pi|]$, $(\beta_i(1), \beta_i(\nu_i))$ est une factorisation de m_i .

Définition 28. Une banque génomique est un ensemble factoriel pour un génome donné.

Définition 29. Soit B une banque génomique. On dit d'un élément de B que c'est un clone.

En génétique expérimentale, le mot clone est à toutes fins pratiques synonyme de copie, c'est-à-dire qu'en fait il existe un procédé largement employé, à savoir la technique d'amplification enzymatique, permettant d'obtenir une multitude de copies d'un segment d'ADN donné. Cela dit, du point de vue axiomatique, dire qu'un segment, au sens mathématique que nous donnons ici, est une copie d'un autre segment ne signifie rien. Par conséquent, de ce point de vue, force est d'admettre que le mot clone est en tout point synonyme du mot segment. Néanmoins, dans la suite, nous n'emploierons le mot clone que pour désigner une séquence appartenant à une banque génomique.

Définition 30. Soit s un segment. On appelle criblage de banque la fonction consistant à faire correspondre à chacun de ses éléments la valeur 0 ou 1 selon que le segment s est un sous-segment de l'élément considéré.

Définition 31. Soit s une séquence d'ADN quelconque. Soit B une banque génomique. On dit alors du type d'opérations expérimentales visant à déterminer, pour chaque clone appartenant à B , si s est un segment de ce clone, que c'est un criblage de B . Plus

précisément, le résultat du criblage de B est la fonction φ_B définie, pour $c \in B$, par

$$\varphi_B(c) = \begin{cases} 1, & \text{si } s \preceq c, \\ 0, & \text{si } s \not\preceq c. \end{cases}$$

Définition 32. Plus généralement, on dit qu'un groupe C de clones est positif par rapport à s s'il existe au moins un clone de C qui soit positif par rapport à s , c'est-à-dire si on a $C \cap R_s(B) \neq \emptyset$. On dit aussi plus simplement, lorsqu'aucun risque de confusion quant au segment s n'est à craindre, que C est positif.

Définition 33. Soit U un ensemble fini et soit R une relation. On note alors $R(U)$ l'ensemble des éléments $u \in U$ tels que u vérifie R , c'est-à-dire qu'on pose

$$R(U) = \{u \in U \mid u \text{ vérifie } R\}.$$

Soit $u \in R(U)$ quelconque. On dit alors de u qu'il est positif par rapport à R , ou, plus simplement, que u est positif, lorsqu'aucun risque de confusion n'est à craindre quant à la relation R .

Définition 34. Soit U un ensemble fini et soit R une relation. Soit $\pi \in \mathcal{P}(U)$. On dit alors que π est positif par rapport à R si π admet au moins un élément positif, c'est-à-dire si $\pi \cap R(U) \neq \emptyset$. On dit aussi plus simplement que π est positif, lorsqu'aucun risque de confusion n'est à craindre.

Définition 35. Soit U un ensemble fini. Soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$. Soit l'application $f_{\mathcal{P}} : \mathcal{P} \rightarrow \{0, 1\}$ définie, pour $\pi \in \mathcal{P}$, par $f(\pi) = 1$ si, et seulement si, π est positif par rapport à R . Plus précisément, on définit $f_{\mathcal{P}}$, pour $\pi \in \mathcal{P}$, par

$$f_{\mathcal{P}}(\pi) = \begin{cases} 1, & \text{si } \pi \cap R(U) \neq \emptyset, \\ 0, & \text{si } \pi \cap R(U) = \emptyset. \end{cases}$$

On dit alors de $f_{\mathcal{P}}$ que c'est l'analyse de \mathcal{P} , relativement à R .

Soit U un ensemble fini, et soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$ quelconque. Notons qu'une analyse de \mathcal{P} permet généralement de déterminer au moins partiellement $R(U)$,

c'est-à-dire permet de déterminer qu'une certaine partie π de $R(U)$ est formée d'éléments positifs, sans toutefois permettre de déterminer qu'ils le sont tous.

Définition 36. Soit U un ensemble fini, et soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$. Soit $u \in U$ quelconque. On note alors $\Phi_{\mathcal{P}}(u)$ l'ensemble des groupes auxquels u appartient. Plus précisément, on a une famille $(\Phi_{\mathcal{P}})_{\mathcal{P} \in \mathcal{R}[U]}$ d'applications définie, pour $\mathcal{P} \in \mathcal{R}[U]$ quelconque, par

$$\begin{aligned} \Phi_{\mathcal{P}} : U &\rightarrow \mathcal{P}[\mathcal{P}], \\ u &\mapsto \{\pi \in \mathcal{P} \mid u \in \pi\}. \end{aligned}$$

Définition 37. Soit U un ensemble fini. Soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$. On note alors $R(\mathcal{P})$ l'ensemble des plans positifs, par rapport à R , de \mathcal{P} .

Définition 38. Soit U un ensemble fini. Soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$. On note alors $Q_R(\mathcal{P})$ l'ensemble des groupes positifs, par rapport à R , de \mathcal{P} , c'est-à-dire l'ensemble défini par

$$Q_R(\mathcal{P}) = \bigcup_{u \in R(U)} \Phi_{\mathcal{P}}(u).$$

Plus précisément, on a l'application

$$\begin{aligned} Q_R : \mathcal{R}[U] &\rightarrow \mathcal{R}[U], \\ \mathcal{P} &\mapsto \{\pi \in \mathcal{P} \mid \pi \cap R(U) \neq \emptyset\}. \end{aligned}$$

Définition 39. Soit U un ensemble fini. Soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$. Soit $u \in U \setminus R(U)$ quelconque. On dit que la valeur de u est déterminée par rapport au triplet (U, R, \mathcal{P}) , ou, plus simplement, lorsqu'aucun risque de confusion n'est à craindre, que cette valeur est déterminée, l'ensemble des groupes de \mathcal{P} auxquels u appartient compte au moins un groupe négatif, autrement dit si $\Phi_{\mathcal{P}}(u) \setminus Q_R(\mathcal{P}) \neq \emptyset$. On dit encore que u est à valeur déterminée. On note alors $NI(\mathcal{P})$ l'ensemble des éléments négatifs dont la valeur est déterminée.

Définition 40. Soit U un ensemble fini. Soit R une relation. Soit $\mathcal{P} \in \mathcal{R}[U]$. Soit $u \in$

$U \setminus \text{ND}(\mathcal{P})$ quelconque. On dit que, par rapport au triplet (U, R, \mathcal{P}) , u est suspect, si tous les groupes de \mathcal{P} auxquels u appartient sont positifs, autrement dit si $\Phi_{\mathcal{P}}(u) \subseteq Q_R(\mathcal{P})$.

Proposition 1.9. – Soit u un élément suspect. S'il existe $\pi \in \Phi_{\mathcal{P}}(u)$ tel que chacun des autres éléments de π est négatif et à valeur déterminée, c'est-à-dire tel que l'on ait $\pi \setminus \{u\} \subseteq \text{NI}(\mathcal{P})$, alors u est positif.

Définition 41. Soit u un élément suspect. Si u appartient à au moins un groupe $\pi \in \mathcal{P}$ vérifiant la relation $\pi \setminus \{u\} \subseteq \text{NI}(\mathcal{P})$, on dit que u est à valeur déterminée. Dans le cas contraire, on dit que la valeur de u demeure indéterminée, ou encore que u est à valeur indéterminée.

Un plan est d'autant plus adéquat qu'il permet de déduire beaucoup d'information sur $R(U)$.

Définition 42. Soit U un ensemble fini de cardinal n . Soit \mathcal{P} un plan par regroupement sur U . Posons $m = |\mathcal{P}|$. Soit $\varphi : \mathcal{P} \xrightarrow{\sim} [m]$ (resp. $\psi : U \rightarrow [n]$) une numérotation des éléments de \mathcal{P} (resp. U). La matrice d'incidence de \mathcal{P} par rapport au couple (φ, ψ) , notée $M_{\mathcal{P}}^{(\varphi, \psi)}$ ou, plus simplement, $M_{\mathcal{P}}$, est la matrice $M = (m_{ij})$ de type (m, n) sur \mathfrak{B} définie, pour $(i, j) \in [m] \times [n]$, par

$$m_{i,j} = \begin{cases} 1, & \text{si } \psi^{-1}(j) \in \varphi^{-1}(i), \\ 0, & \text{sinon.} \end{cases}$$

Exemple 11. Posons $U = \{a, b, c, d, e\}$, $\mathcal{P} = \{\{e, b, c\}, \{b, c\}, \{e, a, b, c\}, \{d\}, \{a, c, d\}, \{e, d\}\}$. Soit φ la numérotation de \mathcal{P} définie par $\{e, b, c\} \mapsto 1$, $\{b, c\} \mapsto 2$, $\{e, a, b, c\} \mapsto 3$, $\{d\} \mapsto 4$, $\{a, c, d\} \mapsto 5$, $\{e, d\} \mapsto 6$, et soit ψ la numérotation de U définie par $e \mapsto 1$, $a \mapsto 2$, $b \mapsto 3$, $c \mapsto 4$, $d \mapsto 5$. On trouvera en la figure (1.1) la matrice d'incidence de \mathcal{P} par rapport à (φ, ψ) .

Par exemple, on apprend à l'examen de cette matrice que l'élément n° 2 appartient aux groupes n° 2, 3 et 5, puisque les termes de la deuxième colonne de $M_{\mathcal{P}}$ correspondant respectivement aux lignes 2, 3 et 5 valent 1.

$$M_{\mathcal{P}}^{(\varphi, \psi)} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Figure 1.1 Matrice d'incidence de \mathcal{P} par rapport au couple (φ, ψ)

Soit R une relation telle que les éléments n° 1 et 2, respectivement, sont positifs, et telle que les autres éléments de U soit négatifs.

On peut alors déduire de \mathcal{P} que l'élément n° 5 est négatif, puisque le groupe n° 4 ne contient que cet élément. On en déduit que l'élément n° 1 est positif, étant donné que le groupe n° 6 est positif, qu'on sait que l'élément n° 5 est négatif, et que l'élément n° 1 est le seul autre élément de ce groupe.

Ce plan permet de déterminer que le fragment n° 5 est négatif, puisque ce fragment est le seul élément du groupe n° 4. Le groupe n° 6 permet alors de déterminer que le fragment n° 1 est positif, puisqu'on sait déjà que le le fragment n° 5 est négatif et que le fragment n° 1 est le seul autre fragment de ce groupe. Par contre, on ne peut déterminer la valeur d'aucun autre élément. On a donc la classification suivante

numéro de colonne	1	2	3	4	5
groupe n° 1	+		-	-	
groupe n° 2		+	-		
groupe n° 3	+	+	-	-	
groupe n° 4					-
groupe n° 5		+		-	-
groupe n° 6	+				-

Dans ce tableau, le signe + indique la présence d'un élément positif dans le groupe

correspondant, le signe – la présence d'un élément négatif, alors que l'absence d'un signe indique naturellement l'absence de l'élément dans le groupe correspondant.

Considérons le plan correspondant à une microplaque formée de 96 cupules (fig. 1.2), comme on en trouve fréquemment dans les laboratoires de génétique, par exemple. Chacune des lignes est un groupe, de même que chacune des colonnes. En fait, le plan est formé des lignes et des colonnes, et d'aucun autre ensemble de cupules, ce qui, en notation mathématique, s'écrit $L \in P$ (resp. $C \in P$) pour L une ligne (resp. une colonne) arbitraire, P désignant le présent plan.¹ Sur les 96 cupules, exactement 2 contiennent un échantillon positif. A priori, l'expérimentateur ne dispose pas de cette information. L'analyse a précisément pour but d'en tirer cette même information, mais en général la valeur de certains échantillons reste indéterminée même après analyse. Le résultat de l'analyse de chacun des groupes est indiqué dans la colonne de droite, pour les lignes, et dans la ligne du bas, pour les colonnes. On voit donc que l'analyse de ce plan permet d'en déduire que tous les échantillons appartenant aux lignes (resp. colonnes) autres que les lignes 3 et 5 (resp. 3 et 8) sont négatifs. Par contre, cette analyse ne permet pas de déterminer la valeur des échantillons appartenant aux diverses intersections de ces deux lignes et de ces deux colonnes. En effet, chacun des éléments correspondants à l'une ou l'autre de ces quatre intersections n'appartient qu'à des groupes positifs, comme on s'en convainc aisément à l'examen du tableau associé (fig. 1.2). Plus précisément, les valeurs respectives des échantillons dont les vecteurs de coordonnées respectifs appartiennent à l'ensemble $\{(3,3), (3,8), (5,3), (5,8)\}$ demeurent indéterminées.

1. Il importe de noter ici que chacune des lignes (resp. colonnes) appartient à P . Ainsi, on se gardera de croire que chacune des lignes ou chacune des colonnes soit incluse dans P , cela n'ayant pas de sens ici, puisque P et une ligne arbitraire L ou une colonne arbitraire C ne sont pas des parties d'un même sur-ensemble.

	1	2	3	4	5	6	7	8	9	10	11	12	
1	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	+	-	-	-	-	-	-	-	-	-	+
4	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	+	-	-	-	-	+
6	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-
	-	-	+	-	-	-	-	+	-	-	-	-	-

Figure 1.2 Tableau des valeurs des échantillons : microplaque de 96 cupules

CHAPITRE II

MÉTHODE SIMPLICIALE

2.1 Introduction

Dans ce chapitre, nous nous intéressons à une classe particulière de plans par regroupement, à savoir la classe des complexes simpliciaux (Park *et al.*, 2003). Nous présentons des résultats, faisant appel à cette classe de plans, grâce auxquels on a des méthodes permettant de déterminer tous les éléments positifs.

Dans le but de clarifier et, en un sens, de simplifier l'article de Park *et al.* il nous a paru opportun de faire appel à la théorie des espèces de structures. Selon nous, cette théorie fournit un langage permettant d'en arriver à une présentation élégante et conceptuellement satisfaisante des concepts qu'on y trouve. De plus, quoi que l'article de Park *et al.* soit clair une fois qu'on a saisi deux ou trois éléments implicites, il faut tout de même réussir à franchir l'obstacle que constitue précisément le fait que ces mêmes éléments soient implicites. En revanche, la théorie des espèces de structures présente l'intérêt d'exprimer en clair ce qui est implicite dans l'article de Park *et al.*.

Dans cette optique, nous introduisons de nouvelles espèces de structures, entre autres non seulement l'espèce des complexes simpliciaux mentionnée plus haut, mais aussi l'espèce des plans d -disjoints, d désignant un nombre naturel donné.

2.2 Préliminaires

En relation avec la notion de graphe simple, on pourra par exemple en guise de complément consulter l'excellent ouvrage de J. Labelle (Labelle, 1981).

Rappelons d'abord ce qu'est un graphe simple.

Définition 1. Soit U un ensemble fini arbitraire. On définit alors l'espèce des graphes simples, notée \mathcal{G} , en posant

$$\mathcal{G}[U] = \{(U, A) \mid A \subseteq \mathcal{P}_2[U]\}.$$

On dit alors d'un élément quelconque $g = (U, A)$ de $\mathcal{G}[U]$ que c'est un graphe sur U . De plus, on dit des éléments de U (resp. A) que ce sont les sommets (resp. les arêtes) de g .

Définition 2. Soit U un ensemble fini quelconque. Soit $g \in \mathcal{G}[U]$. On note alors $A(g)$ l'ensemble des arêtes de g .

Soit U un ensemble fini quelconque. Posons $n = |U|$. Soit $g = (U, A) \in \mathcal{G}[U]$. Soit Π un polyèdre comportant au moins n sommets. Considérons une injection φ de U dans l'ensemble des sommets de Π , telle que, quelle que soit $a = \{u, v\} \in A$, $\varphi(u)$ et $\varphi(v)$ appartiennent à une même arête de Π . On a alors que φ induit une injection $\bar{\varphi}$ de A dans l'ensemble des arêtes de Π . Intuitivement, il est loisible de dire que le couple $(\varphi(U), \bar{\varphi}(A))$ est une « représentation » de g . Les mots « sommet » et « arête » viennent alors de ce que, par rapport à une telle représentation, l'image d'un élément quelconque de U (resp. A) est un sommet de Π (resp. une arête de Π). On a ainsi l'origine intuitive du concept de graphe.

En général, pour raison de commodité, il est évidemment plus simple de se donner une représentation planaire d'un graphe plutôt qu'une représentation spatiale. Dans ce cas, on représente les éléments de l'ensemble U dans le plan, en mettant en bijection U et un ensemble de points géométriques. On représente ensuite les différentes arêtes de g en reliant les points du plan par paire lorsque deux points forment l'image par cette bijection d'un élément de A . Par exemple, posons $U = [5]$. Soit $g \in \mathcal{G}[U]$ le graphe défini

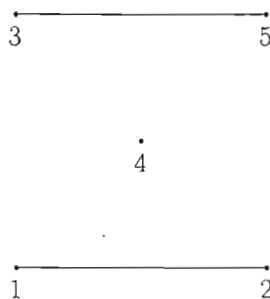


Figure 2.1 Une représentation planaire de g

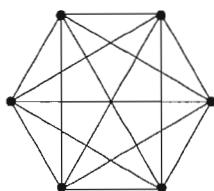


Figure 2.2 Le graphe complet à 6 sommets

en posant $g = (U, A)$, où $A = \{\{1, 2\}, \{3, 5\}\}$. On peut alors voir une représentation planaire de g en la figure 2.1.

Bien entendu, dans le cas d'une représentation planaire, à strictement parler, on fait correspondre à chaque $a \in A$ non pas une arête, au sens usuel du terme, mais plutôt un segment de droite du plan.

Définition 3. Soit U un ensemble fini quelconque. Soient $g, h \in \mathcal{G}[U]$. On définit alors sur $\mathcal{G}[U]$ la relation binaire notée \preceq en posant

$$h \preceq g \iff A(h) \subseteq A(g).$$

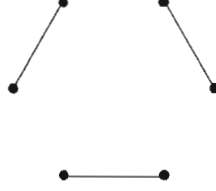


Figure 2.3 Un couplage du graphe complet à 6 sommets

Soit U un ensemble fini quelconque, et soient $g, h \in \mathcal{G}[U]$ tels que $h \preceq g$. On voit donc que la relation \preceq exprime simplement le fait que l'ensemble des arêtes de h est inclus dans l'ensemble des arêtes de g . En particulier, on vérifie immédiatement que cette relation est réflexive, antisymétrique et transitive. Notons que cette relation est plus stricte que la relation « être un sous-graphe », puisque, par définition, les ensembles de sommets respectifs de g et de h coïncident.

Définition 4. Soit U un ensemble fini quelconque. Soit $g = (U, A) \in \mathcal{G}[U]$. Soit $B \subseteq A$. On dit que B est un couplage de g si (U, B) est un sous-graphe de g et si B est formé d'arêtes deux à deux disjointes, c'est-à-dire si, quelles que soient $a, a' \in B$ telles que $a \neq a'$, on a $a \cap a' = \emptyset$.

2.2.1 Graphes et complexes simpliciaux

Comme on le verra, l'espèce des complexes simpliciaux, que nous allons introduire ici, donne lieu à des résultats grâce auxquels on obtient des matrices d -disjointes, concept que nous définirons plus loin.

Définition 5. Soit U un ensemble fini arbitraire. On définit l'espèce \mathcal{S} des complexes simpliciaux, qui est une sous-espèce de l'espèce des plans par regroupement, en posant

$$\mathcal{S}[U] = \{\mathcal{P} \in (\mathcal{P} \circ \mathcal{P})[U] \mid (\forall \pi \in \mathcal{P})(\pi' \subseteq \pi \Rightarrow \pi' \in \mathcal{P})\}.$$

Autrement dit, l'ensemble $\mathcal{S}[U]$ est formé des ensembles \mathcal{P} de parties de U tels que toute partie π appartenant à \mathcal{P} est telle que toute partie π' de π appartient elle-même à \mathcal{P} . On dit alors d'un élément quelconque de $\mathcal{S}[U]$ que c'est un complexe simplicial sur U , ou encore, plus simplement, que c'est un complexe simplicial.

Exemple 1. Soit $U = \{a, b, c, d\}$. Posons

$$\begin{aligned}\mathcal{P}_1 &= \{\{a, b\}, \{a\}, \{b\}, \emptyset\}, \\ \mathcal{P}_2 &= \{\{a, b, c\}, \{a, b\}, \{a, c\}, \emptyset\}.\end{aligned}$$

On vérifie aisément que \mathcal{P}_1 est un complexe simplicial. En revanche, \mathcal{P}_2 n'est pas un complexe simplicial. Pour s'en convaincre, il suffit de noter que le singleton $\{b\}$ n'appartient pas à \mathcal{P}_2 , alors que ce même singleton est inclus dans l'ensemble $\{a, b\}$, lequel appartient à \mathcal{P}_2 .

Définition 6. Soit U un ensemble fini quelconque. Soit \mathcal{S} un complexe simplicial quelconque sur U . On dit alors d'un élément quelconque de U que c'est un sommet de \mathcal{S} . De plus, on dit d'un élément quelconque de \mathcal{S} que c'est une face de \mathcal{S} . Plus précisément, on dit d'un élément de \mathcal{S} de cardinal k , c'est-à-dire formé de k sommets, que c'est une k -face de \mathcal{S} , ou, plus simplement, que c'est une k -face.

Notons que chacune des 1-faces d'un complexe simplicial s'identifie canoniquement à l'unique sommet qu'elle contient.

Exemple 2. Soit T le tétraèdre régulier, et soit U l'ensemble des sommets de T . Posons $U = \{A, B, C, D\}$. Faisons correspondre à chaque sommet de T le singleton de $\mathcal{P}[U]$ formé de ce sommet, à chaque arête de T la paire formée des deux extrémités de cette arête, à chaque face de T l'ensemble de cardinal 3 formé des sommets de cette face, et au tétraèdre lui-même l'ensemble U . On vérifie alors que, par rapport à cette correspondance, la famille Δ de parties de U formée des images respectives par cette correspondance des sommets, des arêtes, des faces de T et de l'ensemble vide est un

complexe simplicial¹. On a en effet

$$\begin{aligned}\Delta = & \{\emptyset, \{A\}, \{B\}, \{C\}, \{D\}, \{A, B\}, \{A, C\}, \{A, D\}, \{B, C\}, \{B, D\}, \{C, D\}, \\ & \{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{B, C, D\}, X\}.\end{aligned}$$

Par exemple, l'élément $\{A, B, C\}$ est donc une 3-face. Chacun des singletons $\{A\}$, $\{B\}$, $\{C\}$, $\{D\}$ peut être vu comme étant un sommet. Du point de vue géométrique, seules les 3-faces, par rapport à la correspondance que nous venons de décrire, peuvent donc s'interpréter comme étant des faces au sens usuel du terme. Par exemple, les 2-faces correspondent à des arêtes.

Définition 7. Soit U un ensemble fini quelconque. Soit R une relation. On dit de R qu'elle est décroissante si, quel que soit $g \in \mathcal{G}[U]$, si g vérifie cette relation, alors quel que soit $h \in \mathcal{G}[U]$ tel que $h \preceq g$, h vérifie lui-même cette relation. Autrement dit, si, quel que soit l'ensemble fini U qu'on considère et quel que soit $g \in \mathcal{G}[U]$, on a l'implication

$$R(g) \implies ((\forall h \in \mathcal{G}[U] \ni (h \preceq g))(R(h))).$$

Définition 8. Soit U un ensemble fini quelconque. Soit R une relation. On dit de R qu'elle est croissante si, quel que soit $g \in \mathcal{G}[U]$, s'il existe un graphe $h \in \mathcal{G}[U]$ tel que $h \preceq g$, vérifiant cette relation, alors g vérifie lui-même cette relation. Autrement dit, si, quel que soit l'ensemble fini U qu'on considère et quel que soit $g \in \mathcal{G}[U]$, on a l'implication

$$(\exists h \in \mathcal{G}[U] \ni ((h \preceq g) \wedge R(h))) \implies R(g).$$

Définition 9. Soit R une relation. On dit de R qu'elle est monotone si elle est croissante ou décroissante.

1. À vrai dire, cela est évident, puisque Δ n'est autre que l'ensemble des parties de X . Néanmoins, l'intérêt de cet exemple réside dans son caractère géométrique.

Définition 10. Soit R une relation.

i) si R est décroissante, on pose

$$\Delta_R^g = \{A(h) \mid (h \preceq g) \wedge R(h)\}.$$

ii) si R est croissante, on pose

$$\Delta_R^g = \{A(h) \mid (h \preceq g) \wedge \neg(R(h))\}.$$

Autrement dit, si par exemple R est décroissante, l'ensemble Δ_R^g est formé des ensembles de la forme $A(h)$, pour h appartenant à l'ensemble des graphes sur U tels que $h \preceq g$ vérifiant R .

Proposition 2.1. – Soit R une relation. Soit U un ensemble fini quelconque. Soit g un graphe quelconque sur U . Si R est monotone, alors Δ_R^g est un complexe simplicial sur $A(g)$, autrement dit, on a $\Delta_R^g \in \mathcal{S}[A(g)]$.

Démonstration. Considérons d'abord le cas où R est décroissante. Soit $\mathcal{A} \in \Delta_R^g$ quelconque. Par définition, l'ensemble \mathcal{A} est de la forme $A(h)$, pour un certain graphe $h \in \mathcal{G}[U]$ vérifiant R . Soit $\mathcal{B} \subseteq \mathcal{A}$ quelconque. Posons $k = (U, \mathcal{B})$. On en déduit, par hypothèse sur R , que k vérifie R , et ainsi on a bien $\mathcal{B} \in \Delta_R^g$.

Dans le cas où R est croissante, on montre aisément que Δ_R^g est un complexe simplicial, par un argument symétrique à celui que nous venons d'invoquer dans le cas où R est décroissante. \square

2.2.2 Monoïdes de matrices booléennes

Comme on le verra, les matrices booléennes servent à coder des ensembles de parties d'un ensemble et des relations d'inclusion entre ensembles. On obtient ainsi des objets mathématiques immédiatement traduisibles en langage machine.

Définition 11. On appelle monoïde un ensemble muni d'une loi de composition associative unifère².

Proposition 2.2. – La loi d'addition booléenne donnée par le tableau suivant

\oplus	0	1
0	0	1
1	1	1

munit l'ensemble $\{0, 1\}$ d'une loi de composition associative unifère. Autrement dit, l'ensemble $\{0, 1\}$ muni de la loi de composition notée \oplus est un monoïde. De plus, cette loi est commutative.

Démonstration. On vérifie aisément que la loi \oplus est associative et commutative et que 0 en est l'élément neutre. \square

Définition 12. On note \mathfrak{B} l'ensemble $\{0, 1\}$ muni de la loi d'addition booléenne. Autrement dit, on pose $\mathfrak{B} = (\{0, 1\}, \oplus)$. Nous appellerons monoïde de Boole ce monoïde.

En particulier, on vérifie très aisément que, quels que soient $n \in \mathbf{N}$ et $x_1, x_2, \dots, x_n \in \mathfrak{B}$, on a l'équivalence

$$x_1 \oplus x_2 \oplus \dots \oplus x_n = 0 \iff x_i = 0, i = 1, 2, \dots, n.$$

Soient $\mu, \nu \in \mathfrak{B}$. Rappelons qu'on désigne par $\mathbf{M}_{\mu, \nu}(\mathfrak{B})$ l'ensemble des matrices de type (μ, ν) sur \mathfrak{B} , c'est-à-dire l'ensemble des matrices sur \mathfrak{B} à μ lignes et ν colonnes.

Définition 13. Soient $\mu, \nu \in \mathbf{N}_+$. Soient $A, B \in \mathbf{M}_{\mu, \nu}(\mathfrak{B})$ quelconques. Posons $A = (a_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq \nu}$, $B = (b_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq \nu}$. On note alors \oplus la loi de composition sur $\mathbf{M}_{\mu, \nu}(\mathfrak{B})$ définie par

$$A \oplus B = (a_{ij} \oplus b_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq \nu}.$$

On dit alors que \oplus est la loi d'addition booléenne définie sur $\mathbf{M}_{\mu, \nu}(\mathfrak{B})$.

2. On dit d'une loi de composition qu'elle est unifère si elle admet un élément neutre.

Autrement dit, on obtient la somme booléenne de deux matrices booléennes appartenant à un même ensemble de la forme $\mathbf{M}_{\mu,\nu}(\mathfrak{B})$ pour certains $\mu, \nu \in \mathbf{N}_+$, en effectuant l'addition terme par terme.

Exemple 3. Posons

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

On a alors

$$A \oplus B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Définition 14. Soient $\mu, \nu \in \mathbf{N}_+$. On note alors $\mathbf{0}_{\mu,\nu}$ la matrice nulle de type (μ, ν) sur \mathfrak{B} . Plus précisément, posons $\mathbf{0}_{\mu,\nu} = (a_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq \nu}$. On définit alors $\mathbf{0}_{\mu,\nu}$ en posant $a_{ij} = 0$, pour $(i, j) \in [\mu] \times [\nu]$.

Proposition 2.3. – Soient $\mu, \nu \in \mathbf{N}_+$. L'ensemble $\mathbf{M}_{\mu,\nu}(\mathfrak{B})$ muni de la loi d'addition booléenne est un monoïde commutatif.

Démonstration. On vérifie aisément que la loi \oplus est associative et commutative et que la matrice $\mathbf{0}_{\mu,\nu}$ en est l'élément neutre. \square

Définition 15. Soient $\mu, \nu \in \mathfrak{B}$ quelconques. On note alors $\mathbf{C}_\mu(\mathfrak{B})$ et $\mathbf{L}_\mu(\mathfrak{B})$ les monoïdes définis en posant

$$\begin{aligned} \mathbf{C}_\mu(\mathfrak{B}) &= \mathbf{M}_{\mu,1}(\mathfrak{B}), \\ \mathbf{L}_\mu(\mathfrak{B}) &= \mathbf{M}_{1,\mu}(\mathfrak{B}). \end{aligned}$$

On dit alors d'un élément quelconque de $\mathbf{C}_\mu(\mathfrak{B})$ (resp. $\mathbf{L}_\mu(\mathfrak{B})$) que c'est un vecteur colonne (resp. un vecteur ligne) de dimension μ sur \mathfrak{B} .

Remarque. Soit $\mu \in \mathbf{N}_+$. Soient $C \in \mathbf{C}_\mu(\mathfrak{B})$, $L \in \mathbf{L}_\mu(\mathfrak{B})$. Soit $i \in [\mu]$ quelconque.

Dans la suite, nous noterons C_i le terme de C correspondant au couple $(i, 1)$. De même, nous noterons L_i le terme de L correspondant au couple $(1, i)$. Plus précisément, posons

$$C = (c_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq 1}, \quad L = (l_{ij})_{1 \leq i \leq 1, 1 \leq j \leq \mu}.$$

Nous posons alors, pour $i \in [\mu]$ quelconque, $C_i = c_{i1}$, $L_i = l_{1i}$.

2.3 Plans d -disjoints

Comme nous l'avons mentionné en introduction, les monoïdes de matrices booléennes permettent de coder des plans par regroupement, de même que des relations d'inclusion entre ensembles, ensembles qu'on peut par exemple considérer en tant qu'éléments de plans par regroupement. Dans cette section, nous allons nous intéresser à une classe particulière de plans par regroupement, à savoir la classe des plans d -disjoints. On utilise le plus souvent le codage à l'aide de matrices booléennes pour coder de tels plans.

Définition 16. Soit U un ensemble fini de cardinal n . Soit φ une bijection de U sur $[n]$. On dit alors de φ que c'est une numérotation de U .

Définition 17. Soit $n \in \mathbb{N}_+$. Soient $C, D \in \mathbf{C}_n(\mathfrak{B})$ quelconques. On définit alors sur $\mathbf{C}_n(\mathfrak{B})$ la relation binaire notée \subseteq en posant

$$C \subseteq D \iff (\forall i \in [n])((D_i = 0) \Rightarrow (C_i = 0)).$$

Exemple 4. Posons

$$C = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

On a alors $C \not\subseteq D$ alors qu'au contraire on a $D \subseteq C$.

Définition 18. Soit U un ensemble fini de cardinal n . Soit φ une numérotation de E .

Soit $\bar{\varphi}$ l'application définie par

$$\begin{aligned}\bar{\varphi}: (\mathcal{P}(E), \cup) &\rightarrow \mathbf{C}_n(\mathfrak{B}), \\ A &\mapsto \bar{\varphi}(A),\end{aligned}$$

où, pour $i \in [n]$, la coordonnée d'indice i de $\bar{\varphi}(A)$ est 1 si, et seulement si, $\varphi^{-1}(i) \in A$. On dit alors de $\bar{\varphi}$ que c'est l'application induite par φ .

Proposition 2.4. — Soit U un ensemble fini de cardinal n . Soit φ une numérotation de U . Soit $\bar{\varphi}$ l'application induite par φ . Alors, $\bar{\varphi}$ est un isomorphisme de monoïdes conservant la relation d'inclusion.

Autrement dit, avec les notations définies dans la définition ci-dessus, on a en particulier que, pour $A, B \in \mathcal{P}[U]$ quelconques, l'application $\bar{\varphi}$ définie plus haut vérifie l'implication

$$(A \subseteq B) \implies (\bar{\varphi}(A) \subseteq \bar{\varphi}(B)).$$

Démonstration. Soit l'application $\bar{\psi} : \mathbf{C}_n(\mathfrak{B}) \rightarrow \mathcal{P}[U]$ définie, pour $C \in \mathbf{C}_n(\mathfrak{B})$ quelconque, par

$$u \in \bar{\psi}(C) \iff C_{\varphi(u)} = 1.$$

On vérifie alors que $\bar{\psi}$ est une bijection. En effet, on vérifie sans trop de peine que $\bar{\psi}$ vérifie les relations

$$\begin{aligned}\bar{\psi} \circ \bar{\varphi} &= \text{Id}_{\mathcal{P}[U]}, \\ \bar{\varphi} \circ \bar{\psi} &= \text{Id}_{\mathbf{C}_n(\mathfrak{B})}.\end{aligned}$$

Par ailleurs, montrons que $\bar{\varphi}$ est un homomorphisme de monoïdes. Soient $A, B \in \mathcal{P}[U]$. Soit $i \in [n]$ quelconque. On a alors, par définition,

$$(\bar{\varphi}(A \cup B))_i = \begin{cases} 1, & \text{si } \varphi^{-1}(i) \in A \cup B, \\ 0, & \text{sinon.} \end{cases}$$

On a donc deux cas à considérer, selon que $\varphi^{-1}(i)$ appartient ou non à $A \cup B$. Dans le premier cas, $\varphi^{-1}(i)$ appartient à l'un au moins des deux ensembles A et B .

Supposons, quitte à intervertir A et B , que $\varphi^{-1}(i)$ appartienne à A . On a alors

$$\begin{aligned} (\bar{\varphi}(A))_i \oplus (\bar{\varphi}(B))_i &= 1 \oplus (\bar{\varphi}(B))_i \\ &= 1 \\ &= (\bar{\varphi}(A \cup B))_i. \end{aligned}$$

Dans le second cas, on a

$$\begin{aligned} (\bar{\varphi}(A))_i \oplus (\bar{\varphi}(B))_i &= 0 \oplus 0 \\ &= 0 \\ &= (\bar{\varphi}(A \cup B))_i. \end{aligned}$$

On a donc montré, i étant quelconque, qu'on a

$$\bar{\varphi}(A \cup B) = \bar{\varphi}(A) \oplus \bar{\varphi}(B),$$

ce qui achève de montrer que $\bar{\varphi}$ est un homomorphisme de monoïdes. \square

L'exemple suivant illustre l'emploi des monoïdes booléens en génomique.

Exemple 5. Soit U l'ensemble formé des trois séquences d'ADN TTCG, ACCT et ACTG. Soit φ la numérotation

$$\begin{aligned} \varphi : \quad U &\rightarrow [3], \\ \text{TTCG} &\mapsto 1, \\ \text{ACCT} &\mapsto 2, \\ \text{ACTG} &\mapsto 3. \end{aligned}$$

Posons $A = \{\text{TTCG}, \text{ACTG}\}$, $B = \{\text{TTCG}, \text{ACCT}\}$. On a alors

$$\bar{\varphi}(A) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \bar{\varphi}(B) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Or, on a évidemment

$$\{\text{TTCG}, \text{ACTG}\} \cup \{\text{TTCG}, \text{ACCT}\} = U.$$

Cette relation correspond alors dans $\mathbf{C}_3(\mathfrak{B})$ à la relation

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \oplus \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

L'exemple suivant illustre la correspondance entre la relation d'inclusion entre deux parties données d'un ensemble et la relation correspondante entre les vecteurs booléens qui leur correspondent.

Exemple 6. Revenons à l'exemple précédent. Posons $C = \bar{\varphi}(\{\text{TTCG}, \text{ACTG}\})$, $D = \bar{\varphi}(\{\text{TTCG}, \text{ACCT}\})$. On a alors $C \not\subseteq D$, puisque $\{\text{TTCG}, \text{ACTG}\} \not\subseteq \{\text{TTCG}, \text{ACCT}\}$. Autrement dit, on a

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \not\subseteq \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Exemple 7. Soit U l'ensemble formé des séquences d'ADN TTCG, ACCT, ACTG et TACG. Soit \mathcal{P} le plan par regroupement sur U défini par

$$\begin{aligned} \mathcal{P} = \{ & \{\text{TTCG}, \text{ACCT}, \text{ACTG}, \text{TACG}\}, \{\text{TTCG}, \text{ACTG}, \text{TACG}\}, \\ & \{\text{TTCG}, \text{ACCT}\}, \{\text{ACCT}\}, \{\text{TACG}\}\}. \end{aligned}$$

Notons que le cardinal de \mathcal{P} vaut 5. Soit φ la numérotation de \mathcal{P} définie par

$$\begin{aligned} \varphi : \quad \mathcal{P} &\rightarrow [5], \\ \{\text{TTCG}, \text{ACCT}, \text{ACTG}, \text{TACG}\} &\mapsto 1, \\ \{\text{TTCG}, \text{ACTG}, \text{TACG}\} &\mapsto 2, \\ \{\text{TTCG}, \text{ACCT}\} &\mapsto 3, \\ \{\text{ACCT}\} &\mapsto 4, \\ \{\text{TACG}\} &\mapsto 5. \end{aligned}$$

Soit ψ la numérotation de U définie par

$$\begin{aligned} \psi : \quad E &\rightarrow [4] \\ \text{TTCG} &\mapsto 1, \\ \text{ACCT} &\mapsto 2, \\ \text{ACTG} &\mapsto 3, \\ \text{TACG} &\mapsto 4. \end{aligned}$$

On a alors que la matrice d'incidence de \mathcal{P} par rapport au couple (φ, ψ) est la matrice

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Définition 19. Soit $\mu, \nu \in \mathbf{N}_+$ et soit $M = (m_{ij})_{1 \leq i \leq \mu, 1 \leq j \leq \nu} \in \mathbf{M}_{\mu, \nu}(\mathfrak{B})$ arbitraire. Soit $j \in [\nu]$. On note alors $C_j(M)$ la j -ième colonne de M . Autrement dit, on pose

$$C_j(M) = \begin{bmatrix} m_{1j} \\ \vdots \\ m_{\mu j} \end{bmatrix}.$$

Définition 20. Soit U un ensemble fini quelconque. Posons $n = |U|$. Soit $d \in \mathbf{N}_+$. On définit alors l'espèce des plans par regroupement d -disjoints, notée $\mathcal{R}^{(d)}$, par

$$\mathcal{R}^{(d)}[U] = \{\mathcal{P} \in \mathcal{R}[U] \mid (\forall \pi \in \mathcal{P})(\forall \Omega \subseteq (\mathcal{P} \setminus \{\pi\}) \ni |\Omega| = d)(\pi \not\subseteq \cup_{\pi' \in \Omega} \pi')\}.$$

Autrement dit, un plan par regroupement \mathcal{P} est un plan d -disjoint si, quel que soit $\pi \in \mathcal{P}$ et quelle que soit la partie Ω de cardinal d de $\mathcal{P} \setminus \{\pi\}$, π n'est pas inclus dans la réunion des éléments de Ω .

En fait, l'image par l'application induite $\bar{\varphi}$ d'un plan d -disjoint est une matrice d -disjointe.

Définition 21. Soient $\mu, \nu \in \mathbf{N}$ et soit $M \in \mathbf{M}_{\mu, \nu}(\mathfrak{B})$. Soit $d \in \mathbf{N}$ arbitraire. On dit que M est d -disjointe si, pour tout $j \in [\nu]$ et pour tout $I \in \mathcal{P}_d([\nu] \setminus \{j\})$, on a

$$C_j(M) \not\subseteq \bigoplus_{i \in I} C_i(M). \quad (2.1)$$

Exemple 8. Montrons que la matrice suivante, considérée évidemment en tant que matrice sur \mathfrak{B} , est 2-disjointe :

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \quad (2.2)$$

Pour ce faire, on doit montrer que, pour $j \in [4]$ et pour tout $I \in \mathcal{P}_2([4] \setminus \{j\})$, la relation (2.1) est satisfaite. Soit $j \in [4]$. Notons d'abord que, pour énumérer les parties à 2 éléments de $[4] \setminus \{j\}$, on peut se ramener à énumérer les suites croissantes de longueur 2 dans $[4] \setminus \{j\}$ (tabl. 2.1). Soit $j \in [4]$ et $I \in \mathcal{P}_2([4] \setminus \{j\})$ quelconques. Posons $I = \{i_1, i_2\}$,

j	$(i_1 < i_2) \subseteq ([4] \setminus \{j\})$
1	(2, 3), (2, 4), (3, 4)
2	(1, 3), (1, 4), (3, 4)
3	(1, 2), (1, 4), (2, 4)
4	(1, 2), (1, 3), (2, 3)

Tableau 2.1 Parties de cardinal 2 de $[4]$

	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$(i_1, i_2) \in [4] \setminus \{j\}$	(2, 3) (2, 4) (3, 4)	(1, 3) (1, 4) (3, 4)	(1, 2) (1, 4) (2, 4)	(1, 2) (1, 3) (2, 3)
i	4 1 2	5 3 2	6 3 1	6 5 4

Tableau 2.2 Indices correspondant aux couples $(j, I) \in \bigcup_{j \in [4]} \{j\} \times \mathcal{P}_2([4] \setminus \{j\})$

avec $i_1 < i_2$. Pour montrer que j et I vérifient la relation (2.1), on peut se ramener à montrer qu'il existe $i \in [6]$ tel que j et I vérifient la relation

$$(C_j(M))_i = 1 \quad \wedge \quad m_{i,i_1} \oplus m_{i,i_2} = 0.$$

On trouvera, en la table (2.2), pour chaque couple $(j, I) \in \bigcup_{j \in [4]} \{j\} \times \mathcal{P}_2([4] \setminus \{j\})$, un indice $i \in [6]$ correspondant tel que cette relation-ci est vérifiée. Nous avons donc de ce fait montré que la matrice (2.2) est 2-disjointe.

Définition 22. Soit $d \in \mathbf{N}_+$. Soient $\mu, \nu \in \mathbf{N}_+$. Soit $M \in \mathbf{M}_{\mu, \nu}(\mathfrak{B})$. On note alors φ_M l'application définie par

$$\begin{aligned} \varphi_M : \mathcal{P}[n] &\rightarrow C_\mu(\mathfrak{B}), \\ I &\mapsto \bigoplus_{j \in I} C_j(M). \end{aligned}$$

Proposition 2.5. – Soit $d \in \mathbf{N}_+$. Soient $\mu, \nu \in \mathbf{N}_+$. Soit $M \in \mathbf{M}_{\mu, \nu}(\mathfrak{B})$. Si M est d -disjointe, alors l'application φ_M est injective.

Démonstration. Soient en effet $I, I' \in \mathcal{P}(n)$, $I \neq I'$. On a alors, M étant d -disjointe,

$$\bigoplus_{j \in I} C_j(M) \neq \bigoplus_{j \in I'} C_j(M).$$

Proposition 2.6. – Soit $d \in \mathbf{N}_+$. Soient $\mu, \nu \in \mathbf{N}_+$. Soit $M \in \mathbf{M}_{\mu, \nu}(\mathcal{B})$. Notons $\bar{\varphi}$ l'application $\bar{\varphi} : \mathcal{P}[n] \rightarrow \text{Im}(\varphi)$ telle que $\bar{\varphi}(\pi) = \varphi(\pi)$, pour tout $\pi \in \mathcal{P}(n)$. Alors, l'application réciproque $\bar{\varphi}^{-1} : \text{Im}(\varphi_M) \rightarrow \mathcal{P}(n)$ est définie, pour $\pi \in \text{Im}(\varphi_M)$ quelconque, par

$$\bar{\varphi}_M^{-1}(C) = \{i \in [n] \mid C_i(M) \subseteq C\}. \quad (2.3)$$

Démonstration. □

Proposition 2.7. – Soit U un ensemble fini. Soit R une relation. Soit \mathcal{P} un plan par regroupement sur U . Soit $Q \subseteq \mathcal{P}$ la partie de \mathcal{P} formée des groupes positifs de \mathcal{P} . Soient φ (resp. ψ) une numérotation de \mathcal{P} (resp. U). Soit $M_{\mathcal{P}} = M_{\mathcal{P}}^{(\varphi, \psi)}$ la matrice d'incidence de \mathcal{P} par rapport au couple (φ, ψ) . Alors, $\varphi(Q) \subseteq \text{Im}(\varphi)$.

Démonstration. Nous allons montrer que $\varphi(Q)$ est l'image par φ_M de $\psi(R(U))$. Il est clair que $Q \subseteq \varphi(\psi(R(U)))$. □

Corollaire. On a $R(U) = \psi^{-1}(\bar{\varphi}_M^{-1}(\varphi(Q)))$.

2.4 Matrices d'incidence associées à un complexe simplicial

Nous allons montrer qu'à partir d'un complexe simplicial donné, on obtient en général plusieurs matrices d -disjointes, pour d appartenant à un certain ensemble d'indices, ensemble dépendant bien entendu du complexe simplicial donné.

Définition 23. Soit U un ensemble fini. Soit $\Delta \in \mathcal{S}[U]$. Soit $d \in \mathbf{N}$. On note alors $F_d(\Delta)$ l'ensemble des d -faces de Δ . Plus précisément, on pose

$$F_d(\Delta) = \{\pi \in \Delta \mid |\pi| = d\}.$$

Définition 24. Soit U un ensemble fini. Soient $k, d \in \mathbf{N}_+$ avec $k > d$. Soit Δ un complexe simplicial sur U tel que $|F_d(\Delta)| = m$ et $|F_k(\Delta)| = n$, c'est-à-dire que Δ compte m d -faces et n k -faces. Soit φ (resp. ψ) une numérotation de $F_d(\Delta)$ (resp. $F_k(\Delta)$). On note alors $M(\Delta, d, k)$ ³ la matrice définie en posant $M(\Delta, d, k) = (m_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$, où

$$m_{ij} = \begin{cases} 1, & \text{si } \varphi^{-1}(i) \subseteq \psi^{-1}(j), \\ 0, & \text{sinon.} \end{cases}$$

On dit alors de $M(\Delta, d, k)$ que c'est la matrice d'incidence de Δ associée au couple (d, k) . À strictement parler, on pourra dire aussi de $M(\Delta, d, k)$ que c'est la matrice d'incidence de Δ associée au triplet $(d, k, (\varphi, \psi))$.

Exemple 9. Soit Δ le complexe simplicial défini en l'exemple 2 de l'article 2.2.1. Rappelons que ce complexe simplicial est construit sur l'ensemble $U = \{A, B, C, D\}$, cet ensemble étant vu comme étant l'ensemble des sommets du tétraèdre régulier. Soit φ la numérotation de $F_2(\Delta)$ définie par

$$\begin{aligned} \{A, B\} &\mapsto 1, \\ \{A, C\} &\mapsto 2, \\ \{A, D\} &\mapsto 3, \\ \{B, C\} &\mapsto 4, \\ \{B, D\} &\mapsto 5, \\ \{C, D\} &\mapsto 6. \end{aligned}$$

Soit ψ la numérotation de $F_3(\Delta)$ définie par

$$\{A, B, C\} \mapsto 1,$$

3. Évidemment, cette matrice dépend aussi du couple (φ, ψ) , encore que deux matrices correspondant respectivement à deux couples distincts s'obtiennent l'une de l'autre par permutation des colonnes et des lignes. Nous omettons le couple (φ, ψ) dans le but de ne pas alourdir indûment la notation.

$$\{A, B, D\} \mapsto 2,$$

$$\{A, C, D\} \mapsto 3,$$

$$\{B, C, D\} \mapsto 4.$$

On vérifie alors que la matrice d'incidence de Δ associée au triplet $(2, 3, (\varphi, \psi))$ est

$$M(\Delta, 2, 3) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad (2.4)$$

Théorème 2.8. – Soit $d \in \mathbf{N}_+$ et soit $k \in \mathbf{N}$ tel que $k > d$. Soit U un ensemble fini. Soit $\Delta \in \mathcal{S}[U]$. Alors, $M(\Delta, d, k)$ est d -disjointe.

Démonstration. Posons $M = M(\Delta, d, k)$. Soit $J \in \mathcal{P}_{d+1}[n]$ quelconque. Soit $j_0 \in J$ quelconque. Posons $J \setminus \{j_0\} = \{j_1, j_2, \dots, j_d\}$. Soit $a_i \in \psi^{-1}(j_0) \setminus \psi^{-1}(j_i)$, $i = 1, 2, \dots, d$, c'est-à-dire que, pour $i \in [d]$, a_i appartient à la k -face numérotée j_0 sans appartenir à la k -face numérotée j_i . Posons $I = \{a_1, a_2, \dots, a_d\}$. Alors, ou bien $|I| = d$, ou bien $|I| < d$. Dans le premier cas, I est une d -face de $\psi^{-1}(j_0)$ mais n'est pas une d -face de $\psi^{-1}(j_i)$, pour $i \in [d]$. Posons $i_0 = \varphi(I)$. On a donc dans ce cas

$$(C_{j_0}(M))_{i_0} = 1 \quad \wedge \quad \bigoplus_{i=1}^d (C_{j_i}(M))_{i_0} = 0,$$

ce qui entraîne immédiatement

$$C_{j_0}(M) \not\subseteq \bigoplus_{i=1}^d C_{j_i}(M).$$

Considérons maintenant le cas où on a $|I| < d$. Soit $J' \subseteq \psi^{-1}(j_0) \setminus I$ tel que $|J'| = d - |I|$. Posons $I' = J' \sqcup I$. Alors, par construction, I' est une d -face de $\psi^{-1}(j_0)$

mais n'est pas une d -face de $\psi^{-1}(j_i)$, pour $i \in [d]$. Posons $i_0 = \varphi(I')$. On a donc là encore

$$(C_{j_0}(M))_{i_0} = 1 \quad \wedge \quad \bigoplus_{i=1}^d (C_{j_i}(M))_{i_0} = 0.$$

Ainsi, dans les deux cas, on a donc $C_{j_0}(M) \not\subseteq \bigoplus_{i=1}^d C_{j_i}(M)$, ce qui achève la démonstration, l'ensemble d'indices J étant quelconque. \square

Soit M une matrice obtenue au moyen du théorème ci-dessus. Une application de ce théorème consiste alors à construire un plan par regroupement de telle sorte que sa matrice d'incidence soit la matrice M . On a alors, puisque M est d -disjointe pour un certain d , une méthode automatique permettant de déterminer l'ensemble des éléments positifs à partir de la donnée des groupes positifs.

Exemple 10. Considérons la matrice (2.4). Supposons qu'il s'agisse là de la matrice d'un plan par regroupement sur un certain ensemble U . Supposons de plus que l'ensemble $R(U)$ des éléments positifs soit l'ensemble $\{\psi^{-1}(4)\}$. On vérifie alors aisément que l'ensemble Q des groupes positifs est l'ensemble $\{4, 5, 6\}$, et ainsi on obtient le vecteur colonne $C = [0, 0, 0, 1, 1, 1]'$. En vertu de la formule (2.3), on devrait avoir $\varphi_M^{-1}(C) = \{4\}$, puisque seule la 4^e colonne est incluse dans C . On voit donc qu'on obtient bien dans cet exemple l'ensemble des éléments positifs, à partir des groupes positifs.

Exemple 11. On vérifie aisément que la matrice (2.4) est 2-disjointe.

Définition 25. Soit U un ensemble fini quelconque. Soit $g \in \mathcal{G}[U]$. Soit $S \subseteq U$. On dit alors de S que c'est un ensemble saturant de sommets si, quel que soit $a \in A(g)$, on a $a \cap S \neq \emptyset$, ou encore si, quel que soit $a \in A(g)$, il existe un sommet u appartenant à S tel que u soit incident à a .

Soit U un ensemble fini quelconque. Soit $g \in \mathcal{G}[U]$. Soit $S \subseteq U$. Notons si besoin est que S est un ensemble saturant de sommets si S vérifie la relation

$$(\forall a \in A(g))(\exists u \in S)(u \in a). \quad (2.5)$$

On utilise le concept d'ensemble saturant de sommets pour en déduire des résultats grâce auxquels on trouve des matrices d -disjointes.

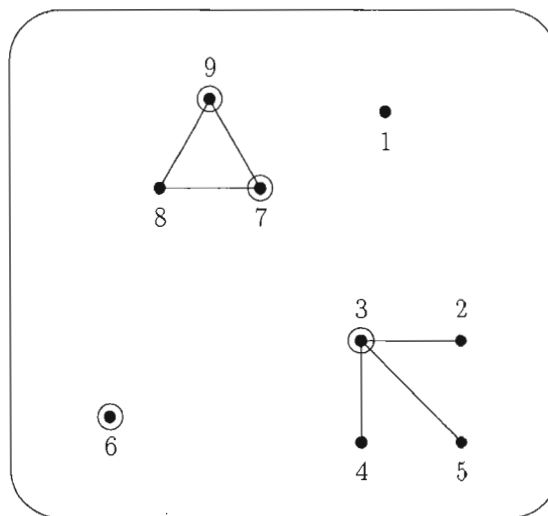


Figure 2.4 Un ensemble saturant de sommets

Exemple 12. Considérons le graphe représenté en la figure (2.4), considéré en tant que graphe sur $[9]$. Soit g ce graphe. Considérons de plus l'ensemble formé des sommets de g que nous avons encerclés. On vérifie alors aisément que S est en fait un ensemble saturant de sommets. Notons d'abord pour ce faire qu'on peut se ramener à considérer le sous-graphe de g dont l'ensemble des sommets est l'ensemble $\{2, 3, 4, 5, 7, 8, 9\}$, puisque les sommets 1 et 6 ne sont incidents à aucune arête. Il y a donc en tout 6 vérifications à faire, puisqu'il y a 6 arêtes dont on doit s'assurer qu'elles vérifient la relation (2.5). Or, on vérifie au premier coup d'oeil que les arêtes $\{3, 2\}$, $\{3, 4\}$ et $\{3, 5\}$ vérifient la relation (2.5), puisque 3 est incident à chacune d'elles, et que 3 appartient à S . Par ailleurs, les arêtes $\{8, 9\}$ et $\{8, 7\}$ vérifient la relation (2.5), puisque 9 est incidente à celle-là et 7 à celle-ci, 9 et 7 appartenant tous deux à S . Finalement, l'arête $\{7, 9\}$ appartient bien à S , puisque, non seulement il existe un sommet u appartenant à S tel que u soit incident à cette arête, mais qu'en plus u n'est pas unique. En effet, de toute évidence, 7 et 9 sont tous deux incidents à cette arête, et, de plus, répétons-le, 7 et 9 appartiennent tous deux à S .

CHAPITRE III

PLANS ORDONNÉS

3.1 Introduction

La classe des plans ordonnés présente entre autres choses l'avantage appréciable de donner souvent lieu à des formules algébriques simples pour les statistiques et les paramètres qu'on leur associe. Nous présentons dans ce chapitre un exposé détaillé portant sur cette classe de plans et sur diverses statistiques qu'on lui associe (Balding *et al.*, 1996; Torney, 1999).

3.2 Plans ordonnés

Remarque. Soient $n \in \mathbb{N}$, F une espèce de structures. Dans la suite, afin d'alléger les notations, nous noterons $F[n]$ l'ensemble $F[[n]]$.

Définition 1. Soit $n \in \mathbb{N}_+$. On note alors $\mathcal{P}_{\times n}$ l'espèce de structures définie par

$$\mathcal{P}_{\times n} = \underbrace{\mathcal{P} \times \cdots \times \mathcal{P}}_{n \text{ facteurs}}.$$

Cette espèce de structures est appelée l'espèce des plans par regroupement ordonnés de longueur n . On dira plus simplement aussi plans ordonnés de longueur n . Plus précisément, pour un ensemble fini arbitraire U , on dira d'un élément de $\mathcal{P}_{\times n}[U]$ que c'est un plan (par regroupement) ordonné de longueur n sur U . Plus simplement, n étant arbitraire, on pourra dire que c'est un plan (par regroupement) ordonné, sans en mentionner la longueur.

Soit U un ensemble fini arbitraire. Soit $n \in \mathbb{N}$. Soit \mathcal{O} un plan par regroupement ordonné de longueur n sur U . On voit donc par là que \mathcal{O} n'est rien d'autre qu'un n -uplet de parties de U . Par suite, à strictement parler, pour un indice $i \in [n]$ quelconque, le groupe \mathcal{O}_i est une coordonnée de \mathcal{O} .

Exemple 13. Posons $U = [4]$, $n = 3$. Posons

$$\mathcal{O} = (\{1\}, \{3, 4\}, \{2\}).$$

On a alors que \mathcal{O} est un plan par regroupement ordonné de longueur 3 sur $[4]$. On écrit alors $\mathcal{O} \in \mathcal{P}_{\times 3}[4]$.

On a une suite (S_1, S_2, \dots, S_v) , où, pour $i \in [v]$, S_i est une suite de la forme (X_1, \dots, X_n) , où X_1, X_2, \dots, X_n sont des variables de Bernoulli deux à deux indépendantes, de paramètre r . Posons $S_i = (X_{i1}, X_{i2}, \dots, X_{in})$. Qui plus est, les variables X_{ij} , $i = 1, \dots, v$, $j = 1, \dots, n$, sont deux à deux indépendantes.

Supposons donnée une numérotation ϕ de l'ensemble sous-jacent E . Par rapport à cette numérotation, on construit aléatoirement le i -ième groupe en posant

$$P_i = \{x \in E \mid \phi(x) \in J_i\}$$

où on a posé

$$J_i = \{i \in [n] \mid x_{ij} = 1\}.$$

Intuitivement, il est loisible d'imaginer qu'on a v urnes numérotées de 1 à v . On effectue nv jets d'une pièce de monnaie biaisée (sauf si $r = 1/2$). À la suite de chacun des n premiers jets, on met ou non dans l'urne n° 1 l'objet correspondant au jet (par exemple, si on en est au troisième jet, on mettra dans cette urne l'objet numéro 3) selon qu'on obtient face ou non, lorsqu'on adopte la convention voulant que « face » soit un succès, par exemple. On procède ainsi pour la deuxième sous-suite de n jets, et ainsi de suite jusqu'à la dernière sous-suite de n jets.

Posons $E = \{a, b, c\}$, et supposons que l'on ait la numération $a \mapsto 1$, $b \mapsto 2$, $c \mapsto 3$.

Par exemple, si on a $v = 3$ et $n = 3$, et si les première, deuxième et troisième suites sont respectivement $(0, 0, 1)$, $(0, 0, 0)$ et $(1, 1, 0)$, alors on obtient la suite $(\{c\}, \emptyset, \{a, b\})$.

À strictement parler, on obtient un plan en prenant l'image de la suite obtenue, c'est-à-dire en formant l'ensemble des coordonnées non vides deux à deux distinctes. Par exemple, posons $n = 4$, $v = 4$. Supposons qu'on obtienne la suite

$$(\{1, 3, 4\}, \emptyset, \{1, 3, 4\}, \{2, 3\}).$$

Alors, à strictement parler, le plan qu'on obtient est l'ensemble

$$\{\{2, 3\}, \{1, 3, 4\}\}.$$

Par ailleurs, peut-être est-il bon de noter qu'un même plan peut provenir de deux suites distinctes. Ce serait le cas par exemple si on obtenait la suite

$$(\{2, 3\}, \{1, 3, 4\}, \emptyset, \emptyset).$$

Définition 2. Soient $U \in \mathbb{B}$, R une relation. On note alors $R(U)$ l'ensemble défini par

$$R(U) = \{u \in U \mid R(u) \text{ est satisfaite}\}.$$

On dit alors des éléments de $R(U)$ qu'ils sont positifs par rapport à R , et des éléments de $U \setminus R(U)$ qu'ils sont négatifs par rapport à R . Plus simplement, nous dirons, lorsqu'aucun risque de confusion ne sera à craindre (notamment lorsque le contexte ne fera mention que d'une seule relation), que les éléments de U sont positifs et que ceux de $U \setminus R(U)$ sont négatifs.

Pratiquement, convenir de dire des éléments d'une partie d'un ensemble donné U qu'ils sont positifs équivaut à se donner une partie de U et à convenir de dire des éléments de cette partie qu'ils sont positifs. Cela dit, la partie de U dont on convient de dire que ses éléments sont positifs est le plus souvent définie au moyen d'une relation.

Définition 3. On définit comme suit la famille d'applications $(N_{U,R,v})_{(U,R,v) \in \mathbb{B} \times \mathbb{R} \times \mathbb{N}_+}$

$$\begin{aligned} N_{U,R,v} : \mathcal{P}_{\times v}[U] &\rightarrow \mathcal{P}(v) \\ \mathcal{O} &\mapsto \{i \in [v] \mid \mathcal{O}_i \subseteq U \setminus R(U)\}. \end{aligned}$$

Afin d'alléger les notations, nous noterons le plus souvent N lorsqu'aucun risque de confusion ne sera à craindre.

Exemple 14. Soit R la relation « x est pair ». Posons $U = [4]$, $v = 3$. De plus, posons

$$\mathcal{O} = (\{1, 2\}, \{3\}, \{2, 4\}).$$

On vérifie aisément que $\mathcal{O} \in \mathcal{P}_{\times 3}[U]$. On a alors $N(\mathcal{O}) = \{2\}$, puisque seule la deuxième coordonnée est formée uniquement d'éléments négatifs, c'est-à-dire ici impairs.

Définition 4. Soit (U, R, v) un triplet quelconque de la catégorie $\mathbb{B} \times \mathbb{R} \times \mathbf{N}^*$. On définit alors comme suit l'application $\tilde{N}_{U,R,v}$:

$$\begin{aligned} \tilde{N}_{U,R,v} : \mathcal{P}_{\times v}[U] &\rightarrow \mathbb{N} \\ \mathcal{O} &\mapsto |\{x \in U \setminus R(U) \mid C(x) \subseteq [v] \setminus N(\mathcal{O})\}|. \end{aligned}$$

Afin d'alléger les notations, nous noterons le plus souvent \tilde{N} lorsqu'aucun risque de confusion ne sera à craindre, notamment lorsque le contexte ne fera mention que d'une seule relation.

Définition 5. On définit comme suit la famille de statistiques $(N_{U,R,v})_{(U,R,v) \in \mathbb{B} \times \mathbb{R} \times \mathbf{N}^*}$

$$\begin{aligned} N_{U,R,v} : \mathcal{P}_{\times v}[U] &\rightarrow \mathbf{N} \\ \mathcal{O} &\mapsto |\{\}\|. \end{aligned}$$

Afin d'alléger les notations, nous noterons le plus souvent \tilde{N} lorsqu'aucun risque de confusion ne sera à craindre, notamment lorsque le contexte ne fera mention que d'une seule relation.

Exemple 15. Soit R la relation « x est pair ». Posons $U = [4]$, $v = 3$. De plus, posons

$$\mathcal{O} = (\{1, 2\}, \{3\}, \{2, 4\}).$$

On vérifie aisément que $\mathcal{O} \in \mathcal{P}_{\times 3}[U]$. Montrons que $\tilde{N}(\mathcal{O}) = 1$. Le nombre 1 est ici indéterminé, puisque la seule coordonnée auquel il appartienne, à savoir la première coordonnée, est positive. Le nombre 3 est déterminé ici, puisqu'il apparaît dans une coordonnée où tous les éléments sont négatifs, à savoir la deuxième coordonnée.

Lemme. Soit $(U, R, v) \in \mathbb{B} \times \mathbb{R} \times \mathbb{N}^*$ arbitraire. Pour tout $\mathcal{O} \in \mathcal{P}_{xv}[U]$, on a

$$\tilde{N}(\mathcal{O}) = 0 \iff \bigcup_{i \in N(\mathcal{O})} \mathcal{O}_i = U \setminus R(U).$$

Démonstration. En effet, supposons qu'il existe \mathcal{O} tel que l'on ait

$$\bigcup_{i \in N(\mathcal{O})} \mathcal{O}_i \subsetneq U \setminus R(U).$$

Soit alors $x \in U \setminus R(U)$ tel que

$$x \notin \bigcup_{i \in N(\mathcal{O})} \mathcal{O}_i.$$

On a alors, par définition, que x est indéterminée, et ainsi $N(\mathcal{O})$ n'est pas vide. Réciproquement, soit \mathcal{O} tel que l'on ait

$$\bigcup_{i \in N(\mathcal{O})} \mathcal{O}_i = U \setminus R(U).$$

Soit x un élément négatif arbitraire. L'identité ci-dessus entraîne alors immédiatement que x appartient à au moins une coordonnée négative, ce qui montre que x est déterminé. \square

Proposition 3.9. – Soit $(U, R, v) \in \mathbb{B} \times \mathbb{R} \times \mathbb{N}^*$ arbitraire. Posons $\pi = |R(U)|$. On a alors

$$P_r(\tilde{N}_{U,R,v} = 0) = \sum_{i=1}^v \binom{v}{i} (1-r)^{i\pi} (1 - (1-r)^\pi)^{v-i} (1 - (1-r)^i)^{n-\pi}. \quad (3.1)$$

Démonstration. Soit E l'événement « $\tilde{N}_{U,R,v} = 0$ ». Notons d'abord que, en vertu du lemme précédent, E est formé de l'ensemble des plans $\mathcal{O} \in \mathcal{P}_{xv}[U]$ tels que tout élément négatif appartient à au moins une coordonnée négative de \mathcal{O} . Autrement dit, on a

$$E = \{\mathcal{O} \in \mathcal{P}_{xv}[U] \mid \text{pour tout } x \in U \setminus R(U), \text{ il existe } i \in [v] \text{ tel que } x \in \mathcal{O}_i\}.$$

Par ailleurs, soit $i \in [v]$ et soit E_i l'ensemble de plans de E comptant exactement i coordonnées négatives. On se convainc alors aisément que la famille $\{E_i\}_{i \in [v]}$ forme une partition de E .

Soit encore une fois $i \in [v]$, et soit $I \in \mathcal{P}_i(v)$ arbitraire. Soit E_I l'ensemble des plans de E dont l'ensemble des indices des coordonnées négatives est I , c'est-à-dire l'ensemble défini par

$$E_I = \{\mathcal{O} \in \mathcal{P}_v[U] \mid N(\mathcal{O}) = I\}.$$

Là encore, on se convainc aisément que la famille $\{E_I\}_{I \in \mathcal{P}_i(v)}$ forme une partition de E_i . On en déduit immédiatement que la famille $\{E_I\}_{I \in \mathcal{P}(v)}$ forme une partition de l'ensemble E . Or, il est bien connu que l'on a, pour tout $i \in [v]$, $|\mathcal{P}_i(v)| = \binom{v}{i}$. Soient alors I, J deux familles d'indices arbitraires dans $[v]$, à cela près qu'elles ont même cardinal. On se convainc alors sans trop de peine que $P_r(E_I) = P_r(E_J)$. On a donc

$$P_r(\tilde{N}_{U,R,v} = 0) = \sum_{i=1}^v \binom{v}{i} P_r(E_{[i]}).$$

Ainsi, on est ramené à montrer qu'on a

$$P_r(E_{[i]}) = (1-r)^{i\pi} (1 - (1-r)^\pi)^{v-i} (1 - (1-r)^i)^{n-\pi}. \quad (3.2)$$

Soit F' l'événement formé des plans $\mathcal{O} \in E_{[i]}$ tels que, pour $\iota \in [i]$, \mathcal{O}_ι est négatif.

Posons $U \setminus R(U) = \{\nu_1, \nu_2, \dots, \nu_{n-\pi}\}$. Posons alors

$$F''_j = \{\mathcal{O} \in E_{[i]} \mid (\nu_j \in \mathcal{O}_1) \vee \dots \vee (\nu_j \in \mathcal{O}_i)\}, \quad j \in [n-\pi].$$

Posons alors

$$F'' = \bigcap_{j \in [n-\pi]} F'_j.$$

On vérifie aisément que $F'' \subseteq F'$. ce qui entraîne, en vertu d'un résultat bien connu de la théorie des probabilités, que $P_r(F' F'') = P_r(F') P_r(F'')$. Or, montrons que $P_r(F') = (1-r)^{i\pi}$. En effet, la probabilité qu'une coordonnée soit négative est égale à $(1-r)^j$, puisque l'événement « la coordonnée ι est négative » est égal à l'événement « la coordonnée ι ne contient aucun des j éléments positifs ».

Soit $j \in [n - \pi]$. Montrons maintenant que $P(F_j'') = 1 - (1 - r)^i$. Or, on a $P_r(F_j'') = 1 - P_r(\mathbb{C}F_j'')$, puisque $P_r(\mathbb{C}F_j'') = (1 - r)^i$. On en déduit, la famille $(F_j'')_{j \in [n - \pi]}$ étant formée d'événements deux à deux indépendants, que $P_r(F'') = (1 - (1 - r)^i)^{n - \pi}$.

Par ailleurs, soit $G \subseteq E_{[i]}$ l'événement formé des plans dont les $v - i$ dernières coordonnées sont positives. Le complément de l'événement « la j -ième coordonnée est positive » étant l'événement « la j -ième coordonnée est négative », on en déduit aisément que $P_r(G_j) = 1 - (1 - r)^j$. On a donc $P_r(G) = (1 - (1 - r)^\pi)^{v - i}$, la famille $(G_j)_{j \in [v] \setminus [i]}$ étant formé d'événements deux à deux indépendants. On a donc finalement

$$\begin{aligned} P_r(E_{[i]}) &= P_r(F') P_r(F'') P_r(G) \\ &= (1 - r)^{i\pi} (1 - (1 - r)^i)^{n - \pi} (1 - (1 - r)^\pi)^{v - i}, \end{aligned}$$

ce qui achève la démonstration. \square

Exemple 16. Posons $n = 2$, $P = \{1\}$, $N = \{2\}$, $v = 2$. On a donc $j = 1$. On a alors l'ensemble fondamental

$$\begin{aligned} E = & \{(\emptyset, \emptyset), (\emptyset, \{1\}), (\emptyset, \{2\}), (\emptyset, \{1, 2\}), \\ & (\{1\}, \emptyset), (\{1\}, \{1\}), (\{1\}, \{2\}), (\{1\}, \{1, 2\}), \\ & (\{2\}, \emptyset), (\{2\}, \{1\}), (\{2\}, \{2\}), (\{2\}, \{1, 2\}), \\ & (\{1, 2\}, \emptyset), (\{1, 2\}, \{1\}), (\{1, 2\}, \{2\}), (\{1, 2\}, \{1, 2\})\} \end{aligned}$$

On vérifie alors sans trop de peine que l'ensemble formé des éléments de E contribuant au calcul de $P_1(\tilde{N} = 0)$ est l'ensemble

$$\{\emptyset, \{2\}\}, \{(\{1\}, \{2\}), (\{2\}, \emptyset), (\{2\}, \{1\}), (\{2\}, \{2\}), (\{2\}, \{1, 2\}), (\{1, 2\}, \{2\})\}.$$

Proposition 3.10. — Soit $v \in \mathbf{N}^*$. Posons $\nu = n - \pi$. Soit $i \in [\nu]$. On a

$$P(\tilde{N} = i) = \sum_{k=0}^v \binom{\nu}{k} (1 - r)^{(i + \pi)k} (1 - (1 - r)^k)^{\nu - i} (1 - (1 - r)^\pi)^{v - k}.$$

Proposition 3.11. – Soit $v \in \mathbb{N}^*$. On a

$$E(\tilde{N}) = (n - \pi) (1 - r(1 - r)^\pi)^v.$$

Démonstration. On a

$$\begin{aligned}
E(\tilde{N}) &= \sum_{i=1}^v i P(\tilde{N} = i) \\
&= \sum_{i=1}^v i \sum_{k=0}^v \binom{v}{k} (1 - r)^{(i+\pi)k} (1 - (1 - r)^k)^{\nu-i} (1 - (1 - r)^\pi)^{v-k} \\
&= \nu \sum_{i=1}^v \binom{\nu-1}{i-1} \sum_{k=0}^v \binom{v}{k} (1 - r)^{(i+\pi)k} (1 - (1 - r)^k)^{\nu-i} (1 - (1 - r)^\pi)^{v-k} \\
&= \nu \sum_{k=0}^v \binom{v}{k} (1 - (1 - r)^\pi)^{v-k} \sum_{i=1}^v \binom{\nu-1}{i-1} (1 - r)^{(i+\pi)k} (1 - (1 - r)^k)^{\nu-i} \\
&= \nu \sum_{k=0}^v \binom{v}{k} (1 - (1 - r)^\pi)^{v-k} (1 - (1 - r)^k)^\nu (1 - r)^{jk} \sum_{i=1}^v \binom{\nu-1}{i-1} \frac{(1 - r)^{ik}}{(1 - (1 - r)^k)^i} \\
&= \nu \sum_{k=0}^v \binom{v}{k} (1 - (1 - r)^\pi)^{v-k} (1 - r)^{jk} (1 - r)^k \\
&= \nu \sum_{k=0}^v \binom{v}{k} (1 - (1 - r)^\pi)^{v-k} (1 - r)^{jk} (1 - r)^k \\
&= \nu (1 - (1 - r)^\pi)^v \sum_{k=0}^v \binom{v}{k} \left(\frac{(1 - r)^{\pi+1}}{1 - (1 - r)^\pi} \right)^k \\
&= \nu (1 - (1 - r)^\pi)^v \left(1 + \frac{(1 - r)^{\pi+1}}{1 - (1 - r)^\pi} \right)^v \\
&= \nu (1 - (1 - r)^\pi + (1 - r)^{\pi+1})^v \\
&= \nu (1 - r(1 - r)^\pi)^v.
\end{aligned}$$

□

Exemple 17. Posons $U = \{1, 2\}$, $R(U) = \{2\}$, $v = 1$. On a alors

$$\begin{aligned}
E(\tilde{N}) &= P(\tilde{N} = 1) \\
&= P(\{\emptyset, \{2\}, \{1, 2\}\})
\end{aligned}$$

$$\begin{aligned}
&= (1-r)^2 + r(1-r) + r^2 \\
&= 1 - r(1-r).
\end{aligned}$$

3.3 Détermination de la probabilité à l'aide de la méthode ensembliste

Nous allons maintenant nous intéresser à une méthode (Torney, 1999) permettant la détermination de la probabilité pour les événements appartenant à une certaine classe d'événements que nous décrirons plus loin. On obtient ainsi une formule faisant intervenir une variable désignant le cardinal de l'ensemble des parties positives. On a donc là une méthode qui contraste avec la méthode qu'on a vue dans la section précédente, dans la mesure où on y considère une formule faisant intervenir une variable désignant non pas le cardinal d'un ensemble de parties, mais bien plutôt le cardinal d'un ensemble ordinaire, à savoir le cardinal de l'ensemble formé des éléments positifs.

Proposition 3.12. – Soit $n \in \mathbf{N}$. Soit $\mathcal{E}_n = (\{0, 1\}^n, P_p)$ l'espace probabilisé défini par $P_p(\beta) = p^{|\beta^{-1}(1)|}(1-p)^{n-|\beta^{-1}(1)|}$, où $p \in (0, 1)$. À strictement parler, on a là une famille $(\mathcal{E}_n^{(p)})_{p \in (0, 1)}$ d'ensembles probabilisés. Soit alors U un ensemble de cardinal n . Soit $\lambda \in \mathbf{N}$. Soit $\varphi : [n] \xrightarrow{\sim} U$. Alors, l'application

$$\begin{aligned}
\alpha_U^{(n, \lambda)} : \quad \mathcal{E}_n^\lambda &\rightarrow \mathcal{P}_{\times \lambda}[U], \\
(\beta_1, \dots, \beta_\lambda) &\mapsto (\varphi(\beta_1^{-1}(1)), \dots, \varphi(\beta_\lambda^{-1}(1))).
\end{aligned}$$

est un isomorphisme d'espèces probabilisés

On écrit alors, par abus de langage, un élément quelconque de l'ensemble \mathcal{E}_n^λ au moyen de la bijection canonique $(\beta_1, \dots, \beta_\lambda) \mapsto (\beta_1^{-1}(1), \dots, \beta_\lambda^{-1}(1))$.

Notons que, quel que soit l'ensemble fini U et quel que soit $\lambda \in \mathbf{N}$, la loi de probabilité induite par cet isomorphisme coïncide avec la loi de probabilité définie sur $\mathcal{P}_{\times \lambda}[U]$, lorsqu'on considère l'espèce $\mathcal{P}_{\times \lambda}$ en tant qu'espèce probabilisée. Rappelons que, pour $\mathcal{O} \in \mathcal{P}_{\times \lambda}[U]$ quelconque, la probabilité de \mathcal{O} est donnée par

$$P(\mathcal{O}) = p^{|\mathcal{O}|} (1-p)^{\lambda n - |\mathcal{O}|},$$

où on a posé $||\mathcal{O}|| = \sum_{h=1}^n |\mathcal{O}_h|$.

Soit $i \in [\lambda]$ quelconque. Soit S une relation appartenant à une classe de relations que nous allons définir dans ce qui suit. L'auteur se propose de déterminer la probabilité d'un événement de la forme

$$\{\mathcal{O} \in \mathcal{P}_{\lambda} U \mid S(\mathcal{O}_i)\},$$

c'est-à-dire un événement formé des plans ordonnés sur U de longueur λ tels que la i -ième coordonnée vérifie la relation S .

En fait, on vérifie aisément que la probabilité d'un tel événement est égale, S étant fixée, à la probabilité de l'événement

$$\{\pi \in \mathcal{P}[U] \mid S(\pi)\},$$

où ici \mathcal{P} désigne l'espèce probabilisé des parties d'un ensemble, de paramètre p . On peut donc sans perte de généralité supposer que $\lambda = 1$.

Notons F_S l'événement ci-dessus. Soit R une relation. Soit $i \in [n]$ quelconque. Soit $N \subseteq U \setminus R(U)$ tel que $|N| = i$. L'auteur présente en fait une formule pour la probabilité de F_S dans le cas où S est une relation de la forme

$$\pi \supseteq N \quad \wedge \quad \pi \subseteq U \setminus R(U),$$

où R est une relation donnée et N est une partie négative de U , c'est-à-dire vérifiant la relation $N \subseteq U \setminus R(U)$. Autrement dit, avec les notations ci-dessus, l'auteur présente une formule donnant la probabilité d'un événement de la forme

$$\{\pi \in \mathcal{P}[U] \mid \pi \supseteq N \quad \wedge \quad \pi \subseteq U \setminus R(U)\}.$$

L'auteur raisonne comme suit.

Notons d'abord qu'on peut sans perte de généralité se ramener, au moyen de l'isomorphisme α_U , au cas où l'événement F_S est en fait l'événement

$$\{J \subseteq [n] \mid S(J)\}.$$

Soit donc R une relation. Soit $\varphi : [n] \xrightarrow{\sim} U$. Soit donc $I \subseteq [n]$, $I \subseteq [n] \setminus \varphi^{-1}(R(U))$.
Il s'agit donc de trouver une formule donnant la probabilité de l'événement noté $E_{I,R}$ défini par

$$E_{I,R} = \{J \subseteq [n] \mid J \supseteq I \quad \wedge \quad J \subseteq [n] \setminus \varphi^{-1}(R(U))\}.$$

Posons

$$E'_{I,R} = \{J \subseteq [n] \mid J \supseteq I \quad \wedge \quad J \setminus I \subseteq [n] \setminus \varphi^{-1}(R(U)) \setminus I\}.$$

On vérifie alors aisément que $P(E_{I,R}) = P(E'_{I,R})$. Posons alors

$$F'_{I,R} = \{J' \subseteq [n] \setminus I \mid J' \subseteq [n] \setminus \varphi^{-1}(R(U))\}.$$

On vérifie alors sans trop de peine qu'on a

$$E'_{I,R} = \bigsqcup_{J' \in F'_{I,R}} \{I \sqcup J'\}.$$

On a donc

$$\begin{aligned} P(E'_{I,R}) &= \sum_{J' \in F'_{I,R}} P(\{I \sqcup J'\}) \\ &= p^i \sum_{J' \in F'_{I,R}} p^{|J'|} (1-p)^{|J'|}. \end{aligned}$$

En effet, quel que soit J' , I et J' sont indépendants.

On est donc ramené à tenter de trouver une formule algébrique pour la somme

$$\sum_{J' \in F'_{I,R}} p^{|J'|} (1-p)^{|J'|}.$$

L'auteur déclare avoir trouvé la formule suivante :

$$P(E_{I,R}) = \sum_{h=i}^n \binom{n-i}{h-i} p^h (1-p)^{n-h} \frac{\binom{\binom{n}{i}-j-1}{\binom{h}{i}-1}}{\binom{\binom{n}{i}-1}{\binom{h}{i}-1}}, \quad (3.3)$$

j désignant le nombre de i -parties positives de $[n]$, c'est-à-dire le cardinal de l'ensemble

$$\{J \subseteq [n] \mid |J| = i \quad \wedge \quad J \cap \varphi^{-1}(R(U)) \neq \emptyset\}.$$

Or, nous pensons que le raisonnement conduisant à cette formule comporte une faille. Considérons l'ensemble noté E_I formé des parties de $[n]$ contenant I , c'est-à-dire l'ensemble $\{J \subseteq [n] \mid J \supseteq I\}$. On a alors

$$P(E_I) = \sum_{h=i}^n \binom{n-i}{h-i} p^h (1-p)^{n-h}.$$

En effet, soit la partition $\Pi = \{\Pi_i\}_{i=0}^n$ définie par

$$\Pi_h = \{\pi \in \mathcal{P}([n] \setminus I) \mid |\pi| = h\}, \quad h \in \llbracket n-i \rrbracket.$$

Il est clair que, pour $\pi \in \Pi_h$, on a $P(\pi) = p^h (1-p)^{n-h}$. De plus, on a, par définition,

$$|\Pi_i| = \binom{n-i}{h-i}.$$

On obtient alors une partition $\Pi' = \{\Pi'_i\}_{i=0}^n$ de $F'_{I,R}$ en posant

$$\Pi'_h = \Pi_h \cap F'_{I,R}.$$

On a donc que, pour $\pi \in \Pi'_h$, $P(\pi) = p^h (1-p)^{n-h}$. Ainsi, on a

$$P(F'_{I,R}) = \sum_{h=i}^n |\Pi'_h| p^h (1-p)^{n-h}.$$

Or, on a évidemment

$$P(E'_{I,R}) = \sum_{h=i}^n |\Pi_h| p^h (1-p)^{n-h} \frac{|\Pi'_h|}{|\Pi_h|}.$$

Ainsi, si on montre que l'identité suivante

$$\frac{|\Pi'_h|}{|\Pi_h|} = \frac{\binom{\binom{n}{i}-j-1}{\binom{h}{i}-1}}{\binom{\binom{n}{i}-1}{\binom{h}{i}-1}}$$

est satisfaite, alors on aura du même coup achevé de démontrer la formule (3.3). Selon nous, voici le raisonnement conduisant à cette formule.

Soit $h \in [n - i] + i$. Soit φ_h l'application définie par

$$\begin{aligned} \varphi_h : \mathcal{P}_h[n] &\xrightarrow{\sim} \mathcal{P}_{\binom{h}{i}-1}[\mathcal{P}[[n] \setminus I]] \\ J &\mapsto \{\pi \subseteq J \mid |\pi| = h \quad \wedge \quad \pi \neq I\} \end{aligned}$$

Posons $E_{I,R}^{(h)} = E_{I,R} \cap \mathcal{P}_h[n]$. On vérifie alors aisément que φ_h est injective. On a donc

$$\frac{|\Pi'_h|}{|\Pi_h|} = \frac{|\varphi_h(E_{I,R}^{(h)})|}{|\varphi_h(\mathcal{P}_h[n])|}.$$

Cela dit, φ_h n'est pas surjective. Plus précisément, on a

$$\begin{aligned} \varphi_h(E_{I,R}^{(h)}) &\subsetneq \mathcal{P}_{\binom{h}{i}-1}[\mathcal{P}[[n] \setminus I] \cap E_{I,R}^{(h)}], \\ \varphi_h(\mathcal{P}_h[n]) &\subsetneq \mathcal{P}_{\binom{h}{i}-1}[\mathcal{P}[[n] \setminus I]]. \end{aligned}$$

Plus précisément, on a en général

$$\frac{|\Pi'_h|}{|\Pi_h|} \neq \frac{|\mathcal{P}_{\binom{h}{i}-1}[\mathcal{P}[[n] \setminus I] \cap E_{I,R}^{(h)}]|}{|\mathcal{P}_{\binom{h}{i}-1}[\mathcal{P}[[n] \setminus I]]|}. \quad (3.4)$$

Or, on a

$$\begin{aligned} |\Pi'_h| &= \binom{\binom{n}{i} - j - 1}{\binom{h}{i} - 1}, \\ |\Pi_h| &= \binom{\binom{n}{i} - 1}{\binom{h}{i} - 1}, \end{aligned}$$

ce qui achève de montrer que la relation (3.3) est fausse, si on admet que la relation (3.4) est satisfaite.

La proposition suivante exprime entre autres le fait que la relation (3.4) est satisfaite pour (h, n) appartenant à une partie dénombrable de l'ensemble

$$\bigsqcup_{n \in \mathbf{N}} [n] \times \{n\}.$$

Proposition 3.13. – Soient $h, n \in \mathbf{N}$. Soit $I \subseteq [n]$. Si $n > 5$, si $h = n - 2$ et si $I = \{1, 2\}$ alors on a

$$\text{Im}(\varphi_{h,n}^{(I)}) \neq \mathcal{P}_{\binom{h}{i}-1}[\mathcal{P}_i[[n] \setminus I]].$$

Démonstration. Posons

$$\mathcal{Q}_{h,n}^{(I)} = \{\pi \in \mathcal{P}_h[n] \mid \pi \supseteq I\}.$$

On vérifie aisément que l'image par $\varphi_{h,n}^{(I)}$ de $\mathcal{Q}_{h,n}^{(I)}$ est formé des éléments \mathcal{J} tels que

$$\left| \bigcup_{\pi \in \mathcal{J}} \pi \cup \{I\} \right| = h.$$

Posons $J = [n - 2]$, $\mathcal{J} = \varphi_{h,n}^{(I)}(J)$. Posons alors

$$\mathcal{K} = (\mathcal{J} \setminus \{\{1, n - 2\}\}) \cup \{\{1, n - 1\}\}.$$

Notons que I n'appartient pas à \mathcal{K} , puisque I n'appartient pas à \mathcal{J} . Montrons d'abord qu'on a

$$\mathcal{K} = \binom{h}{2} - 1. \quad (3.5)$$

Par définition, on a $|\mathcal{J}| = \binom{h}{2} - 1$. Or, on a évidemment $\{1, n - 2\} \in \mathcal{J}$. De plus, on a $\{1, n - 1\} \notin \mathcal{J}$, puisque $n - 1 \notin J$, ce qui achève de montrer la relation (3.5) ci-dessus. Montrons maintenant que le cardinal de la réunion des éléments de \mathcal{K} vaut $n - 1$. En effet, on a

$$\begin{aligned} 1 &\in \{1, 3\} \in \mathcal{K} \\ n - 2 &\in \{2, n - 2\} \in \mathcal{K}, \end{aligned}$$

et ainsi, le cardinal de la réunion des éléments de l'ensemble $\mathcal{J} \setminus \{\{1, n - 2\}\}$ vaut $n - 2$. On a donc, sachant que $n - 1$ n'appartient à aucun des éléments de $\mathcal{J} \setminus \{\{1, n - 2\}\}$, que le cardinal de la réunion des éléments de \mathcal{K} vaut $n - 1$. \square

Exemple 18. Posons $n = 6$, $h = 4$, $I = \{1, 2\}$. Posons

$$\mathcal{K} = \{\{1, 3\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}.$$

On a alors

$$\bigcup_{\pi \in \mathcal{K}} \pi = [5],$$

et ainsi on a bien

$$\left| \bigcup_{\pi \in \mathcal{K}} \pi \right| = [5].$$

De plus, la valeur qu'on obtient dans l'exemple suivant est différente de celle qu'on obtient au moyen de la proposition.

Posons $n = 4$, $i = 2$, $h = 3$, $j = 1$. Soit $A = \{1, 2\}$, $J = \{3, 4\}$. Sauf erreur, on obtient le coefficient de $p^h(1-p)^{n-h}$ en dénombrant les n -uplets $v \in \{0, 1\}^n$ satisfaisant à la condition

$$|v| = 3 \quad \wedge \quad A \subseteq v \quad \wedge \quad \forall w \in J, w \not\subseteq v. \quad (3.6)$$

On devrait donc sauf erreur obtenir

$$P(E_{I,R}) = p^2(1-p)^2 + 2p^3(1-p).$$

Or, la formule (3.3) donne

$$P(E_{I,R}) = p^2(1-p)^2 + \frac{6}{5}p^3(1-p).$$

En fait, nous croyons avoir trouvé une formule pour la probabilité d'un ensemble de la forme $E_{I,R}$. Cette formule fait précisément l'objet de la proposition suivante.

Proposition 3.14. — Soit U un ensemble fini. Soit R une relation. Soit $i \in \mathbf{N}$, $i \leq |U|$. Soit $N \subseteq U \setminus R(U)$ tel que $|N| = i$. Posons $\pi = |R(U)|$. On a alors

$$P(E_{I,R}) = p^i(1-p)^\pi.$$

Comme on peut le constater, cette formule ne dépend pas de la variable notée j dans (3.3), c'est-à-dire que, plus précisément, cette formule-ci ne dépend que de la variable i .

En guise de conclusion, nous dirons que si ce que nous venons de présenter est vrai, alors l'analyse qu'on trouve dans (Torney, 1999) resterait à faire.

3.4 Lois de probabilités associées à l'espèce des plans par regroupement

Proposition 3.15. – L'application canonique

$$\mathcal{P}_{\times\lambda}[U] \rightarrow \mathcal{R}_{\leq\lambda}[U]$$

induit une loi de probabilité sur l'ensemble $\mathcal{R}_{\leq\lambda}[U]$.

Proposition 3.16. – Soient $\lambda, m \in \mathbf{N}$. Soit \mathbf{c} une composition de λ de longueur m . Notons $\omega(\lambda, \mathbf{c})$ le nombre de partitions ordonnées de l'ensemble $[\lambda]$, associées à \mathbf{c} . On a alors

$$\omega(\lambda, \mathbf{c}) = \binom{\lambda}{c_1} \binom{\lambda - c_1}{c_2} \cdots \binom{\lambda - c_1 - \cdots - c_{m-1}}{c_m}. \quad (3.7)$$

Proposition 3.17. – Soit U un ensemble fini de cardinal n . Soit $\mathcal{P} \in \mathcal{R}[U]$. Posons $m = |\mathcal{P}|$. Soit $\lambda \in \mathbf{N}$ tel que $\lambda \geq m$. Soit $\varphi : [m] \xrightarrow{\sim} \mathcal{P}$. Posons $\pi_i = \varphi(i)$, $i \in [m]$. Soit $\mathbf{v} \in \mathbf{N}^m$ défini, pour $i \in [m]$, par $v_i = |\pi_i|$. On a alors

$$P(\mathcal{P}) = \sum_{\mathbf{c} \in C(\lambda, m)} \omega(\lambda, \mathbf{c}) p^{\mathbf{c} \cdot \mathbf{v}} (1 - p)^{\mathbf{c} \cdot \mathbf{v}}. \quad (3.8)$$

Proposition 3.17. – On a

$$E(N_{\leq\lambda}) = E(N).$$

CHAPITRE IV

MISE EN APPLICATION DE L'ESTIMATEUR DE MONTE CARLO DE L'ESPÉRANCE DE LA STATISTIQUE Σ

4.1 Introduction

Les plans hypercubiques forment une classe de plans par regroupement qu'on est naturellement conduit à considérer, dans la mesure où on a alors un estimateur du nombre d'éléments positifs, simple du point de vue automatique, c'est-à-dire non seulement facilement traduisible en langage machine, mais aussi tout à fait efficace de ce point de vue. Nous présentons dans ce chapitre un exposé détaillé d'un article (Barillot, Cohen et Lacroix, 1991) portant sur cette classe de plans. Notons que cet article est principalement motivé par des applications à la génomique des méthodes qui y sont présentées. Pour ce faire, nous introduisons le concept d'espèce probabilisée, qui est un cas particulier du concept d'espèce évaluée (Bergeron, Labelle et Leroux, 1994).

4.2 Notions préliminaires

Soient F, G des espèces de structures. Soit U un ensemble fini quelconque. Rappelons qu'on définit le produit cartésien noté $F \times G$ de F et G en posant

$$(F \times G)[U] = F[U] \times G[U].$$

Définition 1. Soit E un ensemble fini, et soit H une espèce de structures. Soit U un ensemble fini quelconque. On définit alors l'espèce des E -familles de H -structures, notée

H_E , en posant, d'une part,

$$H_E[U] = \{\varphi \mid \varphi : E \rightarrow H[U]\}.$$

Soit de plus V un ensemble fini arbitraire et $\sigma : U \xrightarrow{\sim} V$. On définit alors le transport de structures en posant

$$\begin{aligned} H_E[\sigma] : H_E[U] &\rightarrow H_E[V], \\ \psi &\mapsto H[\sigma] \circ \psi. \end{aligned}$$

Soit σ une permutation de E sur $[n]$. On fait de cette espèce une espèce probabilisée en posant

$$P((s_\iota)_{\iota \in E}) = P((s_{\sigma(1)}, \dots, s_{\sigma(n)})) = \prod_{i=1}^n P(\{s_{\sigma(i)}\}).$$

Définition 2. On dit qu'une espèce de structures valuée F est une espèce de structures probabilisée si, quel que soit l'ensemble fini U qu'on considère, le couple $(F[U], w_U)$ est un ensemble probabilisé.

Proposition 4.1. — Soit F une espèce de structures probabilisée. L'espèce valuée dont l'espèce sous-jacente est l'espèce

$$\underbrace{F \times \dots \times F}_{n \text{ facteurs}}$$

et dont la fonction de valuation est définie par $w(s_1, \dots, s_n) = P(s_1) \cdots P(s_n)$, est une espèce probabilisée.

Définition 3. Soit H une espèce de structures probabilisée, et soit $n \in \mathbb{N}$. On appelle *loi séparable* la loi de probabilité qu'on obtient en posant

$$P(\{\{s_1\}, \dots, \{s_n\}\}) = \prod_{i=1}^n P(\{s_i\}).$$

Soit $n \in \mathbf{N}$. Dans la suite, nous noterons \mathbf{H}_n l'espèce de structures définie par

$$\mathbf{H}_n = \underbrace{H \times \cdots \times H}_{n \text{ facteurs}},$$

où, par abus de langage, chacun des facteurs H apparaissant dans le second membre désigne en fait l'espèce de structures sous-jacente à l'espèce probabilisée H .

Proposition 4.2. – L'espèce de structures \mathbf{H}_n munie de la loi séparable est une espèce probabilisée.

Dans la suite du présent ouvrage, nous noterons \mathbf{H}_n l'espèce \mathbf{H}_n munie de la loi séparable.

Exemple 19. Soit \in l'espèce des éléments. Rappelons que cette espèce de structures est définie par $\in[U] = U$, $U \in \mathbb{B}$. On obtient une espèce probabilisée en posant $w_U = 1/|U|$.

Proposition 4.3. – Soit E un ensemble fini. Posons $n = |E|$. On a alors l'isomorphisme d'espèces probabilisées $\mathbf{H}_E = \mathbf{H}_n$.

Démonstration. Soit σ une bijection de E sur $[n]$. Posons α_U .

$$\begin{aligned} \alpha_U : \mathbf{H}_n &\rightarrow \mathbf{H}_E \\ \varphi &\mapsto \varphi \circ \sigma. \end{aligned}$$

Nous allons montrer que la famille $(\alpha_U)_{U \in \mathbb{F}}$ est un isomorphisme naturel. Soit V un ensemble fini de cardinal n arbitraire, et soit σ une bijection de U sur V . On doit montrer que l'on a l'identité

$$H[\sigma] \circ (\varphi \circ \sigma) = (H[\sigma] \circ \varphi) \circ \sigma.$$

Mais cela est une conséquence immédiate de l'associativité de la composition fonctionnelle. □

Corollaire. Soit $l \in \mathbf{N}_+$ et soit $d \in \mathbf{N}_+$. Posons $E = \mathbf{Z}_l^d$, $n = l^d$. On a alors $\mathbf{H}_E = \mathbf{H}_n$.

Démonstration. En effet, le cardinal de \mathbf{Z}_l^d , considéré en tant qu'ensemble, est l^d □

Exemple 20. Posons $l = 2, d = 2$. Posons $E = \mathbf{Z}_2^2$. On a donc, en vertu du corollaire précédent, que l'espèce \mathbf{H}_E est isomorphe à l'espèce \mathbf{H}_4 . Autrement dit, quel que soit l'ensemble U qu'on considère, on a une bijection faisant correspondre à une fonction de \mathbf{Z}_2^2 dans $H[U]$, un quadraplet de H -structures

$$(s_1, s_2, s_3, s_4).$$

Exemple 21. Si $l = 3$ et $d = 2$, on représente l'ensemble \mathbf{Z}_l^d par un réseau carré du type représenté en la figure (4.1).

Définition 4. Soient U, E des ensembles finis. Soient $l, d \in \mathbf{N}_+$. Posons $E = \mathbf{Z}_l^d$. On dit alors dans ce cas d'un élément quelconque de $\mathbf{H}_E[U]$ que c'est une H -configuration hypercubique de paramètre l, d sur U . De plus, si $H = \in$, nous dirons plus simplement, dans la suite du présent ouvrage, d'un élément de $\mathbf{H}_E[U]$, que c'est une configuration hypercubique de paramètre l, d sur U .

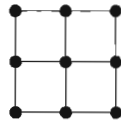


Figure 4.1 Un réseau carré avec $l = 3$

Définition 5. Soit U un ensemble fini. Soit H une espèce de structures. Soit $l, d \in \mathbf{N}_+$. Soit $\gamma \in \mathbf{H}_E[U]$. On associe alors à γ la famille $(\pi_i)_{i \in \llbracket d \rrbracket}$ définie par

$$\begin{aligned} \pi_i(\gamma) : \text{Im}(\gamma) &\rightarrow \llbracket l \rrbracket \\ s &\mapsto \pi_i(\gamma^{-1}(s)), \end{aligned}$$

où, pour $i \in \llbracket d \rrbracket$, π_i désigne la projection sur le i -ième axe de coordonnées.

Définition 6. Soient $l, d \in \mathbf{N}_+$. On note alors $\varphi_{l,d}$ l'application définie par

$$\begin{aligned} \varphi_{l,d}: \mathbf{H}_E[U] &\rightarrow \mathcal{R}[H[U]], \\ \gamma &\mapsto \bigsqcup_{i \in \llbracket d \rrbracket} \bigsqcup_{j \in \llbracket l \rrbracket} \{(\pi_i(\gamma))^{-1}(j)\}. \end{aligned}$$

Définition 7. Soient $l, d \in \mathbf{N}_+$. Posons $E = \mathbf{Z}_l^d$. On note alors $\mathbf{H}_E(l, d)$ l'espèce de structures définie en posant

$$\mathbf{H}_E(l, d)[U] = \text{Im}(\varphi_{l,d})(\mathbf{H}_E[U]).$$

Notons que si $E = \in$, l'espèce $\mathbf{H}_E(l, d)[U]$ est une sous-espèce de l'espèce des plans par regroupement sur U .

4.3 Criblage de banque à l'aide de la méthode de reconfiguration

Soit $n \in \mathbf{N}_+$. Dans ce qui suit, nous allons noter $\Pi(n)$ l'ensemble des puissances strictement positives de n , à savoir l'ensemble défini par

$$\Pi(n) = \{n^k \mid k \in \mathbf{N}_+\}.$$

On note alors Ω l'ensemble des entiers naturels qui sont puissances d'un nombre premier, à savoir l'ensemble défini par

$$\Omega = \bigcup_{p \in \mathfrak{P}} \Pi(p).$$

Voici, dans l'ordre croissant, les nombres ≤ 100 qui sont puissances d'un nombre premier :

$$\begin{aligned} &2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 23, 25, 27, 29, 31, 32, 37, 41, \\ &43, 47, 49, 53, 59, 61, 64, 67, 71, 73, 79, 81, 83, 89, 97, \dots \end{aligned}$$

Soit $n \in \mathbf{N}_+$. Rappelons que l'anneau \mathbf{Z}_n des entiers modulo n est un corps si, et seulement si, n est un nombre premier. Dans ce cas, \mathbf{Z}_n induit donc une structure d'espace vectoriel sur \mathbf{Z}_n^d pour $d \in \mathbf{N}_+$ quelconque. Plus généralement, rappelons aussi (Rotman, 1973) qu'il existe un corps dont l'ordre soit n si, et seulement si, $n \in \mathfrak{Q}$, un tel corps étant unique à isomorphisme près. Soit alors $q \in \mathfrak{Q}$ quelconque. On en déduit donc immédiatement qu'on peut munir l'ensemble $\llbracket q \rrbracket$ d'une structure de corps. Notons alors $\mathcal{G}(q)$ l'ensemble $\llbracket q \rrbracket$ muni d'une structure de corps. On en déduit alors immédiatement que cette structure de corps induit une structure d'espace vectoriel sur l'ensemble $\mathcal{G}(q)^d$, pour $d \in \mathbf{N}_+$ quelconque.

La méthode de reconfiguration consiste d'abord à mettre en bijection l'ensemble des objets qu'on considère et l'ensemble des sommets du réseau hypercubique $V = \mathcal{G}(q)^d$ de dimension d , pour un certain $d \in \mathbf{N}$ et pour un certain $q \in \mathfrak{Q}$. Géométriquement, l'ensemble sous-jacent à l'espace V est l'ensemble des sommets du réseau hypercubique de dimension d et de longueur de côté q . Cela fait, on se donne une matrice inversible (qu'on peut par exemple obtenir aléatoirement). Soit M_1 cette matrice. On applique ensuite M_1 à chacun des vecteurs de coordonnées associées aux objets de la configuration initiale, de façon à obtenir une nouvelle configuration. Si la donnée formée de la configuration initiale et de la nouvelle configuration ainsi obtenue suffit à déterminer tous les objets positifs, alors on a terminé. Sinon on peut itérer ce processus jusqu'à ce qu'on ait réussi cette même détermination. Dans un tel cas, il s'agit de choisir des matrices M_2, M_3, \dots, M_t , pour un certain t , de telle sorte que chaque nouvelle transformation soit la plus différente possible de chacune des précédentes.

Supposons par exemple que le cardinal de l'ensemble d'objets qu'on considère vaille 9. Désignons ces objets par les lettres A à I . Posons $d = 2$ et $n = 3$. On peut alors mettre en bijection cet ensemble et l'ensemble sous-jacent au réseau hypercubique \mathbf{Z}_3^2 . On obtient ainsi une configuration telle que celle représentée en la figure 4.2.

$$\begin{array}{|c|c|c|}
 \hline
 A & B & C \\
 \hline
 D & E & F \\
 \hline
 G & H & I \\
 \hline
 \end{array} . \tag{4.1}$$

Figure 4.2 Configuration hypercubique de paramètres 2, 3

Posons

$$M = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

vue comme matrice à coefficients dans \mathbf{Z}_3 . Notons que M est bien inversible. On obtient alors, lorsqu'on l'applique aux vecteurs de coordonnées respectifs des objets apparaissant dans la configuration ci-dessus, la nouvelle configuration suivante :

$$\begin{array}{|c|c|c|}
 \hline
 A & I & E \\
 \hline
 F & B & G \\
 \hline
 H & D & C \\
 \hline
 \end{array} . \tag{4.2}$$

Figure 4.3 Nouvelle configuration

4.3.1 Condition d'acceptabilité

Soit M une matrices à termes dans un anneau commutatif. Dans ce qui suit, nous allons noter $\mathfrak{m}(M)$ l'ensemble des mineurs de M .

Dans la suite, sauf mention expresse du contraire, chacune des matrices que nous considérerons aura pour ensemble des indices des lignes (resp. colonnes) un ensemble de la forme $\{1, 2, \dots, m\}$. Plus précisément, lorsque nous considérerons une matrice sans préciser quels sont son ensemble des indices des lignes et son ensemble des indices des colonnes, alors on conviendra que l'ensemble $\{1, 2, \dots, m\}$ (resp. $\{1, 2, \dots, n\}$) est son ensemble des indices des lignes (resp. colonnes), où m (resp. n) désigne le nombre de lignes (resp. colonnes) de cette matrice.

Par exemple, si nous posons

$$M = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 1 \end{bmatrix}$$

sans préciser quels en sont les ensembles d'indices, alors on doit comprendre que A a pour ensemble d'indices de lignes l'ensemble $\{1, 2\}$ et pour ensemble d'indices de colonnes l'ensemble $\{1, 2, 3\}$, puisque A comporte 2 lignes et 3 colonnes.

Soit A un anneau commutatif, et soient $m, n \in \mathbf{N}_+$. On note alors $\mathbf{M}_{m,n}(A)$ le module des matrices de type (m, n) à termes dans A . De plus, on note $\mathbf{M}_n(A)$ l'ensemble des matrices à termes dans A dont l'ensemble des indices de colonnes est l'ensemble $\{1, \dots, n\}$. Autrement dit, on pose

$$\mathbf{M}_n(A) = \bigsqcup_{k \geq 1} \mathbf{M}_{k,n}(A).$$

Finalement, dans la suite du présent ouvrage, nous noterons $\mathbf{C}_n(A)$ l'ensemble des matrices carrées à termes dans A dont l'ensemble des indices des lignes et l'ensemble des indices des colonnes est l'ensemble $[n]$.

Définition 8. Soit A un anneau commutatif. Soient $m, n \in \mathbf{N}$, $M \in \mathbf{M}_n(A)$. Soit $H \subseteq [m]$, $K \subseteq [n]$. Posons $M = (m_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$. Soit N l'application définie par

$$\begin{aligned} N : H \times K &\rightarrow A \\ (h, k) &\mapsto m_{hk}. \end{aligned}$$

On dit alors que N est la sous-matrice de M associée au couple d'ensembles (H, K) . On note alors $N = M_{H,K}$.

Exemple 22. Posons

$$M = \begin{bmatrix} 1 & 3 & 2 \\ -2 & 0 & 1 \\ 1 & 1 & 0 \\ 4 & 2 & -3 \end{bmatrix}. \quad (4.3)$$

Par définition, M est une famille indexée par l'ensemble produit $[4] \times [3]$. Autrement dit, on a

$$M = (a_{ij})_{1 \leq i \leq 4, 1 \leq j \leq 3}.$$

Posons $H = \{1, 4\}$, $K = \{1, 2, 3\}$. On a donc

$$H \times K = \{(1, 1), (1, 2), (1, 3), (4, 1), (4, 2), (4, 3)\}.$$

Soit alors N la sous-matrice de M associée au couple (H, K) . On a alors, par abus de langage,

$$N = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 2 & -3 \end{bmatrix}.$$

Remarque. On se gardera de croire que l'ensemble des indices des lignes de N soit l'ensemble $\{1, 2\}$. De même, l'ensemble des indices des colonnes de N n'est pas l'ensemble $\{1, 2, 3\}$. On utilise la représentation ci-dessus par raison de commodité. Intuitivement, comme on peut le voir, la matrice N s'obtient de la matrice M en supprimant de celle-ci la deuxième ligne et la troisième ligne.

Définition 9. Soit A un anneau commutatif, et soit $n \in \mathbf{N}_+$. On note alors \boxplus la loi de composition, appelée loi d'empilement, définie par

$$\begin{aligned} \boxplus : \mathbf{M}_n(A) \times \mathbf{M}_n(A) &\rightarrow \mathbf{M}_n(A) \\ (A, B) &\mapsto A \boxplus B, \end{aligned}$$

où, lorsqu'on pose $M = (m_{ij}) = A \boxplus B$, on a, quel que soit $j \in \{1, 2, \dots, n\}$,

$$m_{ij} = \begin{cases} a_{ij}, & 1 \leq i \leq m, \\ b_{i-m,j}, & m+1 \leq i \leq m+m'. \end{cases}$$

Remarque. On vérifie très aisément que la loi d'empilement n'est pas commutative. En fait, cette loi est analogue à la loi de concaténation.

Intuitivement, on obtient $A \boxplus B$ à partir de A et B tout simplement en « empilant » A par dessus B . De façon informelle, on exprime cela en écrivant

$$A \boxplus B = \begin{bmatrix} A \\ B \end{bmatrix}.$$

Autrement dit, A (resp. B) est la sous-matrice de $A \boxplus B$ dont l'ensemble des couples d'indices est l'ensemble produit $[m] \times [n]$ (resp. $\{m+1, \dots, m+m'\} \times [n]$).

Exemple 23. Posons

$$A = \begin{bmatrix} -1 & 0 & 2 \\ 0 & 3 & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 3 & 4 & -5 \\ 2 & 0 & 1 \\ -1 & 1 & -2 \end{bmatrix}.$$

On a alors

$$A \boxplus B = \begin{bmatrix} -1 & 0 & 2 \\ 0 & 3 & 1 \\ 3 & 4 & -5 \\ 2 & 0 & 1 \\ -1 & 1 & -2 \end{bmatrix}.$$

Définition 10. Soit $p \in \mathbf{N}$ tel que $1 \leq p \leq \min(m, n)$. Soient $H \subseteq [m]$, $K \subseteq [n]$ tels que $|H| = |K| = p$. On dit alors du déterminant de $M_{H,K}$, noté $\det(M_{H,K})$, que c'est un mineur d'ordre p de M

Définition 11. Soit $a \in A$. On dit que a est un mineur de M si, pour un certain $p \in \mathbf{N}$, $p \leq \min(m, n)$, il existe $H \subseteq [m]$, $K \subseteq [n]$, $|H| = |K| = p$, tel que $\det(M_{H,K}) = a$.

Autrement dit, un mineur de M est un mineur d'ordre p de M , pour un certain $p \in [\min(m, n)]$.

Exemple 24. Le déterminant d'une matrice d'ordre 1 étant égal à son unique terme, on en déduit aisément que l'ensemble des mineurs d'ordre 1 d'une matrice est égal à l'ensemble des termes de cette matrice. Par exemple, considérons la matrice

$$M = \begin{bmatrix} 1 & 3 & -2 \\ -2 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \quad (4.4)$$

On a alors que l'ensemble des mineurs d'ordre 1 de M est $\{-2, 0, 1, 3\}$.

Par analogie avec le concept mathématique fondamental d'image, nous dirons que le tableau qu'on obtient en faisant abstraction de l'ensemble des indices des lignes et de l'ensemble des indices des colonnes d'une matrice donnée est l'image de cette matrice. L'image suffit pour obtenir le déterminant, c'est-à-dire que le calcul du déterminant ne fait pas intervenir l'ensemble des indices des lignes et l'ensemble des indices des colonnes.

Posons

$$M = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

On trouvera les mineurs d'ordre 2 de M en le tableau (4.2).

De plus, on a

$$\begin{vmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 3 - 1 + (-1) = 1.$$

On a donc en conclusion que l'ensemble $\mathfrak{m}(M) = \{1, 2, 3, q-1\}$.

Définition 12. Soit $q \in \mathfrak{Q}$. Soit $d \in \mathbb{N}_+$. Soit $M \in \mathbf{C}_d(\mathfrak{S}(q))$. On dit alors que M est acceptable si, quel que soit $\delta \in \mathfrak{m}(M)$, on a $(\delta, q) = 1$, c'est-à-dire si δ et q sont premiers entre eux.

Dans ce qui suit, chaque fois qu'il sera question d'une loi de composition, on

H	K	image	déterminant $\in \mathbf{Z}_q$
$\{1, 2\}$	$\{1, 2\}$	$\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}$	3
$\{1, 2\}$	$\{1, 3\}$	$\begin{smallmatrix} 1 & 1 \\ 2 & 1 \end{smallmatrix}$	$q - 1$
$\{1, 2\}$	$\{2, 3\}$	$\begin{smallmatrix} 1 & 1 \\ 2 & 1 \end{smallmatrix}$	$q - 1$
$\{1, 3\}$	$\{1, 2\}$	$\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}$	1
$\{1, 3\}$	$\{1, 3\}$	$\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}$	3
$\{1, 3\}$	$\{2, 3\}$	$\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}$	1
$\{2, 3\}$	$\{1, 2\}$	$\begin{smallmatrix} 1 & 2 \\ 1 & 1 \end{smallmatrix}$	$q - 1$
$\{2, 3\}$	$\{1, 3\}$	$\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}$	1
$\{2, 3\}$	$\{2, 3\}$	$\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}$	3

Tableau 4.2 Mineurs de M

supposera qu'il s'agit d'une loi associative.

Définition 13. Soit E un ensemble. Soit $I \subseteq \mathbf{N}$. Soit $(e_i)_{i \in I}$ une famille d'éléments de E indexée par I . Soit alors $J \subseteq I$. On note alors $(e_i)_{i \in J}$ la sous-famille de $(e_i)_{i \in I}$, indexée par J .

Définition 14. Soit E un ensemble et soit \perp une loi de composition sur E . On définit alors par récurrence sur $n \in \mathbf{N}$ la notation ci-dessous. Soit $I \subseteq \mathbf{N}$, $|I| < \infty$ et soit $(e_i)_{i \in I}$ une famille d'éléments de E indexée par I . Posons $l = |I|$, et soit $\lambda : [l] \xrightarrow{\sim} I$. On pose alors

$$\underset{i:\lambda}{\perp}(e_i)_{i \in I} = \begin{cases} e_{\lambda_1}, & \text{si } n = 1, \\ (\underset{i:\lambda_{-n}}{\perp}(e_i)_{i \in I \setminus \{\lambda(n)\}}) \perp e_{\lambda_n}, & \text{si } n > 1. \end{cases}$$

où λ_{-n} désigne la sous-suite $(\lambda_1, \dots, \lambda_{n-1})$, c'est-à-dire l'application

$$\begin{aligned} f : [n-1] &\rightarrow \mathbf{N} \\ i &\mapsto \lambda(i). \end{aligned}$$

Considérons d'abord un exemple purement formel.

Exemple 25. Soit E un ensemble et soit \perp une loi de composition définie sur E . Soit $I = \{1, 2, 3\}$ et soit $\lambda = \{1, 3, 2\}$. Soit $(e_i)_{i \in I}$ une famille indexée par I . On a alors

$$\underset{i:\lambda}{\perp}(e_i)_{i \in I} = (\underset{i:\lambda_{-3}}{\perp}(e_i)_{i \in \{1,3\}}) \perp e_2 = (e_1 \perp e_3) \perp e_2 = e_1 \perp e_3 \perp e_2.$$

Exemple 26. Considérons la loi d'empilement, notée \oplus . Soit $\lambda = (1, 3, 2)$. Posons

$$\begin{aligned} M_1 &= \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 2 \\ 2 & 2 & 1 \end{bmatrix}, \\ M_2 &= \begin{bmatrix} 1 & 0 & 2 \end{bmatrix}, \end{aligned}$$

$$M_3 = \begin{bmatrix} 5 & 3 & 4 \\ 3 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Notons au passage qu'on peut supposer que chacune de ces matrices appartient à $\mathbf{C}_3(\mathcal{G}(q))$, pour $q > 5$. Posons $I = \{1, 2, 3\}$. On a alors

$$\begin{aligned} \Xi_{i:\lambda}(M_i)_{i \in I} &= (M_1 \Xi M_3) \Xi M_2 \\ &= \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 2 \\ 2 & 2 & 1 \\ 5 & 3 & 4 \\ 3 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Xi \begin{bmatrix} 1 & 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 2 \\ 2 & 2 & 1 \\ 5 & 3 & 4 \\ 3 & 2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 2 \end{bmatrix}. \end{aligned}$$

Soit $t \in \mathbf{N}$. On note alors $\langle t \rangle$ la suite $(1, 2, \dots, t)$. Par ailleurs, soit λ une suite arbitraire d'entiers. On note alors $\downarrow \lambda \downarrow$ la suite miroir de λ , c'est-à-dire la suite qu'on obtient en parcourant λ de droite à gauche, c'est-à-dire $\downarrow \lambda \downarrow = (\lambda_n, \lambda_{n-1}, \dots, \lambda_1)$.

En fait de suites de matrices de reconfiguration, on ne retient que celles qui satisfont à la condition suivante, dite condition d'acceptabilité.

Définition 15. Soit $q \in \mathfrak{Q}$. Soit $d \in \mathbf{N}_+$. Soit $t \in \mathbf{N}_+$ et soient $M_1, M_2, \dots, M_t \in$

$C_d(\mathcal{G}(q))$. On dit alors que la suite (M_1, M_2, \dots, M_t) est acceptable si la matrice

$$\boxplus_{i:\langle t \rangle} \left(\prod_{j:\downarrow\langle i \rangle} M_j \right) \quad (4.5)$$

est acceptable.

Considérons le cas où $t = 3$. Soient $M_1, M_2, M_3 \in M_n(K)$. Alors, la suite

$$(M_1, M_2, M_3)$$

satisfait à la condition générale d'acceptabilité si la matrice

$$\boxplus_{i:\langle 3 \rangle} \left(\prod_{j:\downarrow\langle i \rangle} M_j \right)$$

est acceptable. Voyons quelle expression on obtient pour cette matrice à l'aide de cette formule-ci. On a

$$\boxplus_{i:\langle 3 \rangle} \left(\prod_{j:\downarrow\langle i \rangle} M_j \right) = M_1 \boxplus (M_2 M_1) \boxplus (M_3 M_2 M_1).$$

À la lueur de cet exemple, on peut donc voir qu'on obtient la matrice (4.5) en empilant successivement les matrices $M_1, M_2 M_1, \dots, M_t M_{t-1} \dots M_1$ dans cet ordre, c'est-à-dire que l'avant-dernière de ces matrices-ci est posée par-dessus la dernière, puis celle qui précède l'avant-dernière est posée par dessus la pile formée de l'avant-dernière et de la dernière, et ainsi de suite jusqu'à la première matrice.

Notons que dans le cas particulier où la suite (M_1, M_2, \dots, M_t) de matrices qu'on considère est constante, c'est-à-dire telle que $M_i = M$, pour $i = 1, 2, \dots, t$, la formule (4.5) se simplifie. On a en effet alors

$$\boxplus_{i:\langle t \rangle} \left(\prod_{j:\downarrow\langle i \rangle} M_j \right) = \boxplus_{i:\langle t \rangle} M^i. \quad (4.6)$$

Autrement dit, on obtient dans ce cas-ci la matrice (4.5) en empilant les puissances successives de M jusqu'à la puissance t , dans l'ordre croissant des exposants.

Définition 16. Soit $q \in \mathfrak{Q}$. Soit $d \in \mathbf{N}_+$. Soit $M \in \mathbf{C}_d(\mathcal{G}(q))$. Soit $t \in \mathbf{N}$. On dit que M est acceptable à l'ordre t si la suite

$$(M, M^2, \dots, M^t)$$

est acceptable. De plus, posons

$$\omega(M) = \max\{k \mid M \text{ est acceptable à l'ordre } k\}.$$

On dit alors de $\omega(M)$ que c'est l'ordre de M .

Exemple 27. Posons

$$M = \begin{bmatrix} 7 & 4 & 1 \\ 3 & 7 & 4 \\ 2 & 3 & 7 \end{bmatrix},$$

considérée comme matrice à termes dans \mathbf{Z}_{43} .

On a alors

$$\begin{aligned} M^2 &= \begin{bmatrix} 20 & 16 & 30 \\ 7 & 30 & 16 \\ 37 & 7 & 20 \end{bmatrix} \\ M^3 &= \begin{bmatrix} 33 & 24 & 36 \\ 42 & 28 & 24 \\ 19 & 42 & 33 \end{bmatrix} \\ M^4 &= \begin{bmatrix} 31 & 21 & 37 \\ 39 & 6 & 21 \\ 24 & 39 & 31 \end{bmatrix} \end{aligned}$$

On vérifie que la matrice $\bigoplus_{i:\langle 3 \rangle} M^i$ est acceptable, alors que la matrice $\bigoplus_{i:\langle 4 \rangle} M^i$ ne l'est pas. On voit donc que l'ordre de M vaut 3.

4.3.2 Exemple de reconfiguration

Par exemple, considérons l'exemple suivant, où on a $d = 2$ et $n = 7$. Considérons la matrice

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

vue comme matrice à coefficients dans \mathbf{Z}_7 . Par exemple, posons $M_2 = A^2$. On a donc

$$M_2 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$$

En appliquant la matrice A_2 aux vecteurs de coordonnées des objets apparaissant dans la deuxième configuration, on obtient une troisième configuration (fig. 4.6).

Considérons le cas où $n = 7$. Dans ce cas, n est bien un nombre premier. Considérons alors la matrice A donnée plus haut. On vérifie aisément que $\det(A) = 1$. De plus, les sous-déterminants sont donnés par les termes de cette matrice, et sont tous différents de 0. L'autre condition est évidente, étant donné que $\det(A) \neq 7$ et que chacun des sous-déterminants de A est lui-même différent de 7. On vérifie aisément que la matrice A^2 satisfait elle aussi à ces mêmes conditions. Il s'agit de déterminer le plus grand nombre possible d'objets positifs, en éliminant à chaque nouvelle étape un certain nombre de faux positifs de l'étape précédente, sans en ajouter. Après un certain nombre de ces étapes, au mieux on a éliminé tous les faux positifs, et il ne reste plus que les objets positifs. Considérons par exemple les trois configurations ci-dessous. Après la première analyse, on aboutit à un certain nombre d'éléments éventuellement positifs, indiqués par le signe $+$. Les objets positifs y sont représentés en caractères gras. Au terme de la deuxième analyse, on élimine un certain nombre de faux positifs à l'aide de l'information provenant de l'analyse précédente. Par exemple, en tenant compte que de la seule deuxième analyse, on ne peut pas conclure que l'objet de vecteur $(4, 6)$ soit négatif. En revanche, on sait déjà, par la première analyse, que cet objet est négatif. De ce fait, on l'élimine de l'ensemble des objets éventuellement positifs. Nous avons indiqué ces objets par le signe $-$.

$(0,0)^+$	$(0,1)^+$	$(0,2)^+$	$(0,3)^+$	$(0,4)$	$(0,5)^+$	$(0,6)$
$(1,0)^+$	$(1,1)^+$	$(1,2)^+$	$(1,3)^+$	$(1,4)$	$(1,5)^+$	$(1,6)$
$(2,0)^+$	$(2,1)^+$	$(2,2)^+$	$(2,3)^+$	$(2,4)$	$(2,5)^+$	$(2,6)$
$(3,0)^+$	$(3,1)^+$	$(3,2)^+$	$(3,3)^+$	$(3,4)$	$(3,5)^+$	$(3,6)$
$(4,0)^+$	$(4,1)^+$	$(4,2)^+$	$(4,3)^+$	$(4,4)$	$(4,5)^+$	$(4,6)$
$(5,0)$	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
$(6,0)$	$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

(4.7)

Figure 4.4 Configuration initiale

$(0,0)^+$	$(6,2)$	$(5,4)$	$(4,6)^+$	$(3,1)$	$(2,3)^+$	$(1,5)^+$
$(1,6)$	$(0,1)$	$(6,3)$	$(5,5)$	$(4,0)$	$(3,2)$	$(2,4)$
$(2,5)^+$	$(1,0)$	$(0,2)$	$(6,4)^+$	$(5,6)$	$(4,1)^+$	$(3,3)^+$
$(3,4)$	$(2,6)$	$(1,1)$	$(0,3)$	$(6,5)$	$(5,0)$	$(4,2)$
$(4,3)^+$	$(3,5)$	$(2,0)$	$(1,2)^+$	$(0,4)$	$(6,6)^+$	$(5,1)^+$
$(5,2)$	$(4,4)$	$(2,6)$	$(2,1)$	$(1,3)$	$(0,5)$	$(6,0)$
$(6,1)$	$(5,3)$	$(4,5)$	$(3,0)$	$(2,2)$	$(1,4)$	$(0,6)$

(4.8)

Figure 4.5 Deuxième configuration

$(0,0)^+$	$(4,5)^-$	$(1,3)^-$	$(5,1)$	$(2,6)$	$(6,4)$	$(3,2)$
$(2,4)$	$(6,2)$	$(3,0)$	$(0,5)$	$(4,3)$	$(1,1)$	$(5,6)$
$(4,1)^+$	$(1,6)^-$	$(5,4)^-$	$(2,2)$	$(6,0)$	$(3,5)$	$(0,3)$
$(6,5)^-$	$(3,3)^+$	$(0,1)^-$	$(4,6)$	$(1,4)$	$(5,2)$	$(2,0)$
$(1,2)^+$	$(5,0)^-$	$(2,5)^+$	$(6,3)$	$(3,1)$	$(0,6)$	$(4,6)$
$(3,6)$	$(0,4)$	$(4,2)$	$(1,0)$	$(5,5)$	$(2,3)$	$(6,1)$
$(5,3)$	$(2,1)$	$(6,6)$	$(3,4)$	$(0,2)$	$(4,0)$	$(1,5)$

(4.9)

Figure 4.6 Troisième configuration

Il est naturel de se demander dans quelle mesure, dans la pratique, ces configurations sont réalisées. Par exemple, dans le cas bi-dimensionnel, c'est-à-dire dans le cas où $d = 2$, a-t-on vraiment une configuration physique formée d'objets physiques disposés aux sommets d'un réseau à coordonnées entières¹ ? En fait, la réponse est oui. Nous nous sommes alors demandé dans quelle mesure une telle méthode est réalisable lorsque par exemple la longueur de côté est 47, comme cela arrive parfois. En effet, on imagine mal devoir permuter 2709 objets. Non seulement une telle tâche est-elle fastidieuse, mais en plus le risque d'erreur est très grand. En fait, la réponse à cette objection est fort simple : on utilise pour cela un robot programmé.

4.4 Les grandes étapes

Soit $d \in \mathbf{N}_+$, $l \in \mathbf{Q}$. Posons $n = l^d$. Soit $p \in (0, 1)$. Soit \mathbf{Z}_l le corps des classes de congruence modulo l . On note alors \mathbf{Z}_l^d l'espace vectoriel de dimension d , produit

1. Dans la vie de tous les jours, un emballage d'oeufs par 6 ou par 12 est un exemple de configuration physique dans laquelle les objets sont disposés aux sommets d'un réseau rectangulaire à coordonnées entières.

cartésien de d copies de \mathbf{Z}_l . Autrement dit, par définition, on a

$$\mathbf{Z}_l^d = \underbrace{\mathbf{Z}_l \times \cdots \times \mathbf{Z}_l}_{n \text{ facteurs}}.$$

Dans ce qui suit, nous aurons l'occasion de tirer au hasard des éléments de l'ensemble sous-jacent à \mathbf{Z}_l^d . On aura probablement compris que, en pratique, du point de vue de la mise en application du présent algorithme, on tire au hasard des vecteurs pris dans l'ensemble

$$\llbracket l-1 \rrbracket \times \cdots \times \llbracket l-1 \rrbracket,$$

en ayant recours pour ce faire à la bijection usuelle entre l'ensemble \mathbf{Z}_l et l'ensemble $\llbracket l-1 \rrbracket$. Plus généralement, afin d'alléger la notation, nous désignerons une classe de congruence précise par le nombre correspondant à cette classe par cette même bijection canonique. Il faudra donc interpréter la signification d'un nombre selon le contexte. Par exemple, le nombre qu'on exprime au moyen du chiffre « 2 » ne désigne pas une même classe de congruence selon que ce chiffre désigne la classe de congruence modulo 4 ou la classe de congruence modulo 5, par exemple. En effet, il est évident que l'on a

$$\{2, 6, 10, 14, \dots\} \neq \{2, 7, 12, 17, \dots\}.$$

Voici les grandes étapes de l'algorithme :

- 1 Produire une valeur x d'une variable $X : B(n, p)$.
- 2 Produire un échantillon aléatoire sans remise de taille x , pris dans l'ensemble \mathbf{Z}_l^d .
- 3 Calculer la moyenne échantillonnale de la statistique Σ .

Soient $n \in \mathbf{N}$, $i \in \llbracket n \rrbracket$. Posons

$$\pi_{n,i} = \binom{n}{i} p^i (1-p)^{n-i}.$$

Il est bien connu que, quel que soit $n \in \mathbf{N}$, la suite $(\pi_{n,i})_{i \in \llbracket n \rrbracket}$ est unimodale. Soit $n \in \mathbf{N}$. Soit

$$M = \max_{i \in \mathbf{N}} \pi_{n,i}.$$

Soit de plus $k \in \llbracket n \rrbracket$. Posons alors

$$\Pi_{k,n} = 1 - \sum_{i=0}^{k-1} \pi_{n,i}.$$

Posons

$$l_i = \log \left(\binom{n}{i} p^i (1-p)^{n-i} \right), \quad i = 0, 1, \dots, n.$$

On a alors

$$\begin{aligned} \log \left(\sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \right) &= \log \left(\sum_{i=0}^k \exp \left(\log \left(\binom{n}{i} p^i (1-p)^{n-i} \right) \right) \right) \\ &= \log \left(\sum_{i=0}^k \exp(l_i) \right) \\ &= \log \left(\exp(L_k) \sum_{i=0}^k \exp(l_i - L_k) \right) \\ &= L_k + \log \left(\sum_{i=0}^k \exp(l_i - L_k) \right). \end{aligned}$$

Soit $k \in \mathbb{N}$. Posons

$$\begin{aligned} S_{k,n} &= \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \\ \Lambda_{k,n} &= -\log(S_{k,n}). \end{aligned}$$

Posons

$$\pi_{n,i} = \binom{n}{i} p^i (1-p)^{n-i}.$$

Il est bien connu que, quel que soit $n \in \mathbb{N}$, la suite $(\pi_{n,i})_{i \in [n]}$ est unimodale. Soit $n \in \mathbb{N}$. Soit

$$M = \max_{i \in \mathbb{N}} \pi_{n,i}.$$

On a alors évidemment

$$\Lambda_{k,n} = \begin{cases} -\left(l_k + \log(\sum_{i=0}^k \exp(l_i - l_k)) \right), & \text{si } k < M, \\ -\left(M + \log(\sum_{i=0}^k \exp(l_i - M)) \right), & \text{si } k \geq M. \end{cases}$$

Ces formules présentent l'avantage de minimiser les exposants de l'argument de la fonction logarithmique.

4.5 Approximations par arrondi à droite de la loi binomiale

Soit X une variable aléatoire de loi binomiale, de paramètres n, p . Dans la suite, nous noterons $\pi_{n,i}$ la probabilité de l'événement $X = i$, quel que soit $i \in \llbracket n \rrbracket$. Autrement dit, nous posons

$$\pi_{n,i} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i \in \llbracket n \rrbracket.$$

Définition 17. Soient $n, p \in \mathbf{N}$. On définit une famille de lois de probabilité indexée par l'ensemble $\llbracket n \rrbracket$ comme suit. Soit $k \in \llbracket n \rrbracket$. La loi de probabilité notée $\text{Bin}^{(k)}(n, p)$, appelée loi binomiale arrondie à droite du rang k , est la loi définie en posant

$$P(X = i) = \begin{cases} \pi_{n,i}, & \text{si } i < k, \\ 1 - \Pi_{k,n}, & \text{si } i = k. \end{cases}$$

Soient $d \in \mathbf{N}$, $q \in \mathfrak{Q}$. Une configuration de type (d, l) est une application

$$\varphi : \mathbf{Z}_l^d \rightarrow \{0, 1\}.$$

On appelle une telle configuration configuration hypercubique de type (d, l) . Si $d = 2$ (resp. $d = 3$), on dit qu'on a une configuration carrée (resp. cubique).

Notons $H(l, d)$ l'ensemble des configurations hypercubiques de type (l, d) . Munissons $H(l, d)$ de la loi de probabilité

$$\begin{aligned} P : H(l, d) &\rightarrow \mathbf{R}, \\ \gamma &\mapsto p^{\Pi(\gamma)} (1-p)^{n-\Pi(\gamma)}. \end{aligned}$$

Nous appellerons *cellules* les éléments de $H(l, d)$. Intuitivement, on fixe un ordre de parcours i_1, i_2, \dots, i_n des cellules de $H(l, d)$. Ensuite, on effectue successivement n essais de Bernoulli de paramètre p , deux à deux indépendants. Selon que la valeur de rang i qu'on obtient est 0 ou 1, on considère que dans la cellule n° i on va ranger un objet

positif ou un objet négatif. Considérons la statistique suivante sur $H(l, d)$, notée Σ , qui fait correspondre à une configuration le nombre d'objets suspects de celle-ci. Autrement dit, on a

$$\begin{aligned} \Sigma : H(l, d) &\rightarrow \mathbf{N} \\ \gamma &\mapsto \Sigma(\gamma), \end{aligned}$$

où $\Sigma(\gamma)$ désigne le nombre d'objets suspects de γ .

Considérons maintenant une famille de parties de $H(l, d)$. Soit $n = l^d$, $k \in \llbracket n \rrbracket$. On note alors $H_{\leq k}(l, d)$ l'ensemble des configurations de type (l, d) comptant au plus k objets positifs. Plus précisément, on pose

$$H_{\leq k}(l, d) = \{\gamma \in H(l, d) \mid \Pi(\gamma) \leq k\}.$$

Notons si besoin est que l'on a la suite d'inclusions

$$H_{\leq 0} \subseteq H_{\leq 1} \subseteq \dots \subseteq H_{\leq n}.$$

Dans la suite du présent ouvrage, nous noterons $\alpha_k : H_{\leq k} \rightarrow \mathbf{R}$ l'application qu'on définit en posant

$$\alpha_k(\gamma) = \begin{cases} p^{\Pi(\gamma)}(1-p)^{n-\Pi(\gamma)}, & \text{si } \Pi(\gamma) \leq k, \\ \Pi_{k,n} / \binom{n}{k}, & \text{si } \Pi(\gamma) = k. \end{cases}$$

Proposition 4.4. – Le couple $(H_{\leq k}, \alpha_k)$ est un espace probablisé.

Démonstration. Montrons que l'on a

$$\sum_{\gamma \in H_{\leq k}} \alpha_k(\gamma) = 1.$$

Pour ce faire, notons que le cardinal de l'ensemble des configurations $\gamma \in H_{\leq k}$ telles $\Pi(\gamma) = k$ est évidemment $\binom{n}{k}$. On a alors

$$\sum_{\gamma \in H_{\leq k}} \alpha_k(\gamma) = \sum_{\gamma \in H_{\leq k-1}} \alpha_k(\gamma) + \sum_{\gamma \in H_k(l, d)} \alpha_k(\gamma)$$

$$\begin{aligned}
&= \sum_{i=0}^{k-1} \pi_{n,i} + \binom{n}{k} \frac{\Pi_{k,n}}{\binom{n}{k}} \\
&= \sum_{i=0}^{k-1} \pi_{n,i} + \Pi_{k,n} \\
&= 1.
\end{aligned}$$

□

Soit $k \in \mathbf{N}$. Dans la suite, nous noterons Σ'_k la restriction à l'espace $H_{\leq k}$ de la statistique Σ . À strictement parler, on doit écrire $\Sigma_{l,d}$. Toutefois, pour alléger les notations, dans ce qui suit, nous omettrons les indices l et d .

Proposition 4.5. – Soit $k \in \mathbf{N}$. Soit $l \in \mathfrak{Q}$, $d \in \mathbf{N}_+$. Soit $\Sigma \equiv \Sigma_{l,d}$ et soit $\Sigma'_k \equiv \Sigma'_{k,l,d}$. Considérons Σ'_k en tant qu'estimateur de $E[\Sigma]$. De ce point de vue, on a alors

$$B(\Sigma'_k) \leq 2n\Pi_{n,k},$$

où $B(\Sigma'_k)$ désigne le biais de Σ'_k , considérée en tant qu'estimateur de $E[\Sigma]$. Autrement dit, on a

$$|E[\Sigma] - E[\Sigma'_k]| \leq 2n\Pi_{n,k}.$$

Démonstration. Posons d'abord $m(k) = \min\{i \mid \sigma_i \geq k\}$, $K = |\text{Im}(\varphi)|$. On a alors, par définition, d'une part,

$$E[\Sigma] = \sum_{i=0}^K \sigma_i P(\Sigma = \sigma_i)$$

Soit $i \geq m(k)$. Posons

$$H_{<k}(\sigma_i) = \{\gamma \in H_{<k} \mid \Sigma(\gamma) = \sigma_i\},$$

$$H_k(\sigma_i) = \{\gamma \in H_k \mid \Sigma(\gamma) = \sigma_i\},$$

$$H_{>k}(\sigma_i) = \{\gamma \in H_{>k} \mid \Sigma(\gamma) = \sigma_i\}.$$

On vérifie aisément que ces trois ensembles sont deux à deux disjoints. De plus, on a

$$\Sigma^{-1}(\sigma_i) = H_{<k}(\sigma_i) \sqcup H_k(\sigma_i) \sqcup H_{>k}(\sigma_i).$$

On en déduit immédiatement

$$P(\Sigma = \sigma_i) = P(H_{<k}(\sigma_i)) + P(H_k(\sigma_i)) + P(H_{>k}(\sigma_i)).$$

On a donc

$$E[\Sigma] = \sum_{i=0}^{m(k)-1} \sigma_i P(\Sigma = \sigma_i) + \sum_{i=m(k)}^K \sigma_i P(\Sigma = \sigma_i) \quad (4.10)$$

$$\begin{aligned} &= \sum_{i=0}^{m(k)-1} \sigma_i P(\Sigma = \sigma_i) + \\ &\quad + \sum_{i=m(k)}^K \sigma_i (P(H_{<k}(\sigma_i)) + P(H_k(\sigma_i)) + P(H_{>k}(\sigma_i))) \end{aligned} \quad (4.11)$$

D'autre part, on a

$$E[\Sigma'_k] = \sum_{i=0}^K \sigma_i P(\Sigma'_k = \sigma_i) \quad (4.12)$$

$$= \sum_{i=0}^{m(k)-1} \sigma_i P(\Sigma'_k = \sigma_i) + \sum_{i=m(k)}^K \sigma_i (\alpha_k(H_{<k}(\sigma_i)) + \alpha_k(H_k(\sigma_i))) \quad (4.13)$$

On déduit des identités (4.11) et (4.13) que l'on a

$$\begin{aligned} E[\Sigma] - E[\Sigma'_k] &= \sum_{i=m(k)}^K \sigma_i (P(H_{<k}(\sigma_i)) + P(H_k(\sigma_i)) + P(H_{>k}(\sigma_i)) - (\alpha_k(H_{<k}(\sigma_i)) + \alpha_k(H_k(\sigma_i)))) \\ &= \sum_{i=m(k)}^K \sigma_i (P(H_k(\sigma_i)) + P(H_{>k}(\sigma_i)) - \alpha_k(H_k(\sigma_i))). \end{aligned}$$

On a donc

$$\begin{aligned} |E[\Sigma] - E[\Sigma'_k]| &= \left| \sum_{i=m(k)}^K \sigma_i (P(H_k(\sigma_i)) + P(H_{>k}(\sigma_i)) - \alpha_k(H_k(\sigma_i))) \right| \\ &= \left| \sum_{i=m(k)}^K \sigma_i P(H_k(\sigma_i)) + \sum_{i=m(k)}^K P(H_{>k}(\sigma_i)) - \sum_{i=m(k)}^K \alpha_k(H_k(\sigma_i)) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=m(k)}^K \sigma_i P(H_k(\sigma_i)) + \sum_{i=m(k)}^K P(H_{>k}(\sigma_i)) + \sum_{i=m(k)}^K \alpha_k(H_k(\sigma_i)) \\
&\leq n \left(\sum_{i=m(k)}^K P(H_k(\sigma_i)) + \sum_{i=m(k)}^K P(H_{>k}(\sigma_i)) + \sum_{i=m(k)}^K \alpha_k(H_k(\sigma_i)) \right) \\
&\leq n \left(\binom{n}{k} p^k (1-p)^{n-k} + \Pi_{n,k+1} + \Pi_{n,k} \right) \\
&= 2n \Pi_{n,k}.
\end{aligned}$$

□

Proposition 4.6. – Soit $l \in \mathfrak{Q}$, $d \in \mathbb{N}_+$. On a

$$\Sigma(H(l, d)) = \{\nu_1^{\alpha_1} \cdots \nu_k^{\alpha_k} \mid \alpha_1 + \cdots + \alpha_k = \delta \wedge \delta \leq d \wedge \nu_1 < \nu_2 < \cdots < \nu_k \leq l\}.$$

Démonstration. Cela résulte de ce que tout élément de l'ensemble $\Sigma(H(l, d))$ s'obtient en faisant un produit de la forme $\nu_1 \cdots \nu_d$, avec $\nu_i \leq l$, $i \in [d]$. □

Exemple 28. Considérons le cas où $d = 2$ et $l = 3$. On a alors

$$\Sigma(H(3, 2)) = \{0, 1, 2, 3, 4, 6, 9\}.$$

En effet, les produits de longueur 1 sont 1, 2, 3. Les produits de longueur 2 sont $2^2 = 4$, $2 \cdot 3 = 6$ et $3^2 = 9$. On a alors

$$E[\Sigma] = \varphi_1(p) + 2\varphi_2(p) + 3\varphi_3(p) + 4\varphi_4(p) + 6\varphi_5(p) + 9\varphi_6(p),$$

où

$$\begin{aligned}
\varphi_1(p) &= p(1-p)^8, \\
\varphi_2(p) &= 18p^2(1-p)^7, \\
\varphi_3(p) &= 6p^3(1-p)^6, \\
\varphi_4(p) &= 18p^2(1-p)^7 + 18p^3(1-p)^6 + 9p^4(1-p)^5, \\
\varphi_5(p) &= 12p^3(1-p)^6 + 24p^4(1-p)^5 + 12p^5(1-p)^4 + 6p^6(1-p)^3,
\end{aligned}$$

$$\begin{aligned}\varphi_6(p) &= 6p^3(1-p)^6 + 36p^4(1-p)^5 + 63p^5(1-p)^4 + 78p^6(1-p)^3 + \\ &\quad + 27p^7(1-p)^2 + 9p^8(1-p) + p^9.\end{aligned}$$

D'autre part, considérons le cas où $k = 7$. On a alors en fait

$$E[\Sigma'_7] = \varphi_1(p) + 2\varphi_2(p) + 3\varphi_3(p) + 4\varphi_4(p) + 6\varphi_5(p) + 9\varphi'_6(p),$$

où

$$\varphi'_6(p) = 6p^3(1-p)^6 + 36p^4(1-p)^5 + 63p^5(1-p)^4 + 78p^6(1-p)^3 + \frac{27}{36}\Pi_{9,7}.$$

4.6 Maximisation du domaine d'évaluabilité

Définition 18. Soient $\mathcal{L}_1, \mathcal{L}_2 \subseteq X^*$. Posons $\mathcal{L}_i = (X_i, V_i, R_i, A_i)$, $i = 1, 2$. Soit $x \in X$. Le produit d'emboîtement de \mathcal{L}_1 par \mathcal{L}_2 , noté $\mathcal{L}_1 \sqsubset_x \mathcal{L}_2$, est le langage $\mathcal{L} = (X, V, R, A)$ qu'on obtient en posant

$$\mathcal{L} = (X, V, R, A_1),$$

où

$$\begin{aligned}X &= X_1 \cup X_2, \\ V &= V_1 \cup V_2, \\ R &= R_1 \cup R_2 \cup R_{x, \mathcal{L}_1, \mathcal{L}_2},\end{aligned}$$

où $R_{x, \mathcal{L}_1, \mathcal{L}_2} = \sigma_2|_{x \rightarrow A_2}(R_1)$.

Autrement dit, le produit d'emboîtement définit une loi de composition

$$\begin{aligned}\sqsubset_x : X^* \times X^* &\rightarrow X^* \\ (\mathcal{L}_1, \mathcal{L}_2) &\mapsto \mathcal{L}_1 \sqsubset_x \mathcal{L}_2.\end{aligned}$$

Dans la définition suivante, peut-être est-il bon de souligner que le mot exp est un lexème, autrement dit est un terminal, et non un mot du langage qu'on y considère.

De même, il faut comprendre que les symboles « (» et «) » (parenthèse ouvrante et parenthèse fermante) sont eux aussi des terminaux.

Notons \mathcal{L} le langage des expressions arithmétiques, auquel on adjoint les lexèmes exp et log . Dans ce qui suit, nous allons considérer un « petit » langage, évidemment, comme on s'en convaincra aisément, inclus (strictement) dans \mathcal{L} .

Définition 19. Soit \mathcal{E} le langage formé des sommes formelles finies d'expressions de la forme $\text{exp}(x)$, où x désigne un nombre réel. Plus précisément, \mathcal{E} désigne le langage $\mathcal{L}_{\subset x\mathcal{F}}$, \mathcal{L} désignant le langage $\mathcal{L} = \{X, V, R, A\}$, où

$$\begin{aligned} X &= \{\text{exp}, (,), +\}, x \\ V &= \{A, T, F\}, \\ R &= \{A \rightarrow T, \\ &\quad T \rightarrow \text{exp}(F) \mid T + T\}. \end{aligned}$$

et où \mathcal{F} désigne le langage formé des nombres de type « virgule flottante »².

Exemple 29. Montrons que l'expression

$$\text{exp}(-2) + \text{exp}(19) + \text{exp}(57)$$

appartient bien au langage \mathcal{E} . On a la dérivation

$$\begin{aligned} A &\rightarrow T \\ &\rightarrow T + T \\ &\rightarrow T + T + T \\ &\rightarrow \text{exp}(-2) + T + T \\ &\rightarrow \text{exp}(-2) + \text{exp}(19) + T \\ &\rightarrow \text{exp}(-2) + \text{exp}(19) + \text{exp}(57). \end{aligned}$$

Dans la suite, nous utiliserons le symbole Σ dans son sens usuel en mathématique. Il est bien entendu, dans le contexte présent, ce signe appartient au métalangage, et non

2. Nous entendons par là les nombres de type *float* en C++.

au langage. De même, dans ce qui, nous utiliserons le signe \dots . Là encore, il est bien entendu que ce signe appartient lui aussi au métalangage.

Définition 20. Soit $S \in \mathcal{E}$. Posons $S = \sum_{i=1}^n \exp(x_i)$. Une mise en évidence de S est un couple $(T, S') \in \mathcal{E} \times \mathcal{E}$ tel que T est de la forme $\exp(x)$, satisfaisant à $x + y_i = x_i$, $i = 1, 2, \dots, n$.

Exemple 30. Posons

$$S = \exp(0, 011) + \exp(2, 3) + \exp(4).$$

Alors, le couple (T, S') où $T = \exp(0, 011)$ et $S' = \exp(0) + \exp(2, 289) + \exp(3, 989)$ est une mise en évidence.

Définition 21. Soit $S \in \mathcal{E}$. On dit que le couple (T, S') est une mise en évidence distinguée de S si (T, S') est une mise en évidence de S et si $T = \exp(M)$, où $M = \max x_i$. Cette mise en évidence étant unique, nous dirons que c'est la mise en évidence distinguée de S .

Définition 22. Soit $\mathcal{L} = (X, V, R, A)$ un langage. Soit $r \in R$ de la forme (A, m) . On note alors \mathcal{L}_r le sous-langage de \mathcal{L} formé des expressions qu'on obtient au moyen des dérivations de la forme (u_0, u_1, \dots, u_n) , $n \in \mathbb{N}$, avec $u_0 = (A, m)$.

Exemple 31. Considérons le langage \mathcal{L} défini plus haut. Posons $r = (A, \log(F))$. Alors, on se convainc sans peine que le langage \mathcal{L}_r n'est autre que l'ensemble des expressions de la forme $\log(\cdot)$.

Définition 23. Soit \mathcal{L} un langage formé d'expressions arithmétiques, $m \in \mathcal{L}$. On dit que m est évaluable si $\text{év}(m) \neq \infty$.

Proposition 4.7. – Soit $m \in \mathcal{L}_r$. Soit $D = D(m)$ la factorisation distinguée de m . Alors, D est σ -évaluable.

Démonstration. Posons

$$D = \exp(M) (\exp(x_1) + \dots + \exp(x_n)).$$

On a alors évidemment

$$D \xrightarrow{d} M(\exp(x_1) + \cdots + \exp(x_n)).$$

Or, par définition, on a $x_i \leq 0$, $i = 1, 2, \dots, n$. On a donc

$$\text{év}[(\exp(x_1) + \cdots + \exp(x_n))] \leq n,$$

ce qui entraîne évidemment

$$\text{év}[(\exp(x_1) + \cdots + \exp(x_n))] < \infty.$$

Posons

$$\text{év}[(\exp(x_1) + \cdots + \exp(x_n))] = x.$$

On a donc

$$\text{év}[M(\exp(x_1) + \cdots + \exp(x_n))] = M \cdot x < \infty.$$

□

Soit $M = \min\{i \in \mathbf{N} \mid \text{év}(\exp(i)) < \infty\}^3$.

Exemple 32. Supposons pour fixer les idées que l'on ait $M = 800$. Posons

$$S = \exp(1701) + \exp(900) + \exp(800).$$

On a donc

$$D(S) = \exp(1701)(\exp(0) + \exp(-801) + \exp(-901)).$$

On a alors

$$D(S) \xrightarrow{d} 1600(\exp(0) + \exp(-700) + \exp(-800)) \xrightarrow{\text{év}} 1600(1 + 0 + 0) = 1600 < \infty.$$

3. Par exemple, dans le cas du compilateur CodeWarrior sous MacOS X, $M = 800$

Par contre, considérons la mise en évidence suivante de S :

$$S \equiv \exp(900)(\exp(801) + \exp(0) + \exp(-100)).$$

On a

$$\exp(900)(\exp(801) + \exp(0) + \exp(-100)) \xrightarrow{\text{év}} \infty(\infty + 1 + \exp(-100)) = \infty \cdot \infty = \infty,$$

et ainsi cette expression-ci n'est pas évaluable.

Par abus de langage, nous dirons qu'une somme S est évaluable si elle admet une mise en évidence qui soit évaluable.

Corollaire. Soit $S \in \mathcal{L}$. Si $nM(S) < M$, alors S est évaluable.

En pratique, la condition ci-dessus revient à la condition $M(S) < M'$. En effet, sauf dans des cas improbables, le nombre de termes est très petit devant $M(S)$. On voit donc par là que la transformation logarithmique permet d'étendre prodigieusement le domaine d'évaluabilité de sommes de probabilités binomiales.

Notons que, du point de vue axiomatique, dans le présent contexte, le mot *expression* est en fait synonyme de *mot*, au sens de la théorie des langages.

Définition 24. Soit $\mathbf{u} = (u_0, u_1, \dots, u_k)$ une dérivation. Soit $\mathcal{L}' \subseteq \mathcal{L}$ l'ensemble des expressions de L de dérivation initiale

Définition 25. Soit $\mathbf{u} = (u_0, u_1, \dots, u_k)$ une dérivation avec u_0 de la forme $A \rightarrow m$. On note alors $\varphi_{\mathbf{u}}$ la fonction

$$\begin{aligned} \varphi_{\mathbf{u}}: \mathcal{L} &\rightarrow \mathcal{L} \\ m &\mapsto m_{\mathbf{u}}, \end{aligned}$$

définie comme suit. Soit $m \in \mathcal{L}$ et soit $(u'_0, u'_1, \dots, u'_n)$ tel que $S \xrightarrow{*} m$. On pose alors que $m_{\mathbf{u}}$ satisfait à $S \xrightarrow{n} m_{\mathbf{u}}$ pour $m'_{\mathbf{u}}$.

Proposition 4.8. – Soit $\mathcal{L}' \subseteq \mathcal{L}$ le langage formé des mots ayant en commun la dérivation (u_0, u_1, u_2) , où $u_0 = (A \rightarrow \log(F))$, $u_1 = (F \rightarrow FF)$, $u_2 = (F \rightarrow \exp(x))$,

$u_3 = (F \rightarrow \exp(x))$. Ce langage est formé des expressions de la forme $\log(\exp(x)\exp(x))$, où $x \in \mathcal{F}$. Considérons la dérivation (u'_0, u'_1, u'_2, u'_3) où $u'_0 = [A \rightarrow F]$, $u'_1 = [F \rightarrow FF]$, $u'_2 = [F \rightarrow x]$, $u'_3 = [F \rightarrow \exp(x)]$. Soit \mathcal{L}'' le langage formé des mots de la forme $x\exp(x)$. Alors, ce langage n'est autre que l'image par $\varphi_{u'}$ de \mathcal{L}' .

On peut s'en convaincre à l'aide de l'exemple suivant. Posons $m = \log(\exp(2)\exp(5))$.

On a

$$\begin{aligned} A &\rightarrow \log(F) \\ &\rightarrow \log(FF) \\ &\rightarrow \log(\exp(x)\exp(x)) \\ &\rightarrow \log(\exp(2)\exp(5)). \end{aligned}$$

D'autre part, on a

$$\begin{aligned} A &\rightarrow F \\ &\rightarrow FF \\ &\rightarrow x\exp(x) \\ &\rightarrow 2\exp(5). \end{aligned}$$

CHAPITRE V

SIMULATIONS

5.1 Introduction

Nous donnons ici des tableaux de simulations de l'estimateur de Monte Carlo que nous avons décrit dans le chapitre précédent. Plus précisément, pour chaque couple de paramètre (l, d) et pour p donné, on trouve dans ces tables l'espérance du nombre de faux positifs pour les plans hypercubiques de paramètres l, d et de loi de probabilité de paramètre p . Il est intéressant de noter que, par simulation, on peut obtenir non seulement le nombre d'éléments suspects d'un plan aléatoire donné, mais aussi le nombre d'éléments positifs. On obtient alors le nombre de faux positifs en retranchant du nombre d'éléments suspects le nombre d'éléments positifs. Bien entendu, en génétique expérimentale, dans le cas d'un plan non adaptatif, on n'est pas en mesure en général de déterminer le nombre de faux positifs d'un plan donné.

5.2 Tables

Dans tous les tableaux que nous donnons ici, à savoir les tableaux 5.1 à 5.5, le nombre d'itérations vaut 100. Plus précisément, le nombre d'itérations (on dit encore le nombre de *répétitions*) de l'estimateur de Monte Carlo est tout simplement le nombre de termes de la somme apparaissant dans la formule

$$\frac{1}{n} \sum_{i=1}^n f_i,$$

où f_i désigne le nombre de faux positifs du plan produit lors de la i -ième itération, pour des valeurs données des paramètres l et d et p . On devrait donc plus précisément écrire

$$\frac{1}{n} \sum_{i=1}^n f_{i,l,d}.$$

Ainsi, on produit successivement 100 plans hypercubiques aléatoires de paramètres l , d et p donnés, et à chaque itération on détermine le nombre de faux positifs. Si on additionne les 100 nombres qu'on obtient ainsi, et qu'on divise ensuite la somme obtenue par 100, on obtient alors une estimation de l'espérance du nombre de faux positifs associée aux plans hypercubiques de paramètres l , d et p . Notons $F_{l,d}$ l'estimation ainsi obtenue, associée aux paramètres l , d et p . À l'examen de ces tables, on constate que, pour p fixé et pour l croissant, $F_{l,d}$ croît, ce qui en fait selon nous est tout à fait naturel.

$l \setminus p$	0,001	0,002	0,004	0,005	0,01	0,02	0,03	0,05	0,07	0,1	0,2
2	0	0	0	0	0	0	0,02	0	0	0,06	0,22
3	0	0	0	0,02	0,04	0,03	0,03	0,1	0,23	0,35	1,13
4	0	0	0	0	0,02	0,06	0,16	0,42	0,56	1,09	3,13
5	0	0,02	0,04	0,06	0,07	0,18	0,4	0,8	1,9	2,84	6,74
7	0,02	0,04	0,08	0,08	0,18	0,69	1,22	2,98	6,32	9,91	21,52
8	0,02	0,06	0,07	0,08	0,37	1,07	2,56	6,46	10,28	16,98	32,28
9	0,04	0,05	0,06	0,16	0,55	2,01	4,19	9,5	13,54	25,01	47,52
11	0,06	0,07	0,2	0,29	1,05	4,01	8,31	18,97	28,96	47,57	77,51
13	0,05	0,08	0,52	0,6	2,76	7,31	14,35	34,06	52,92	80,98	120,47
16	0,07	0,24	0,9	1,31	4,87	15,51	32,65	73,98	105,85	150,51	191,95
17	0,06	0,29	1,21	1,83	6,48	21,44	42,97	90,36	125,91	173,64	218,35
19	0,12	0,52	1,69	2,96	10,56	33,46	66,33	123,97	174,44	234,55	280,71
23	0,28	0,98	3,81	5,24	17,75	67,69	126,55	230,03	313,34	386,16	419,79
25	0,38	1,25	4,96	8,58	27,9	91,06	159,4	303,44	395,24	473,79	497,83
27	0,54	1,8	7,92	11,08	37,56	131,7	220,25	384,49	488,07	568,43	580,7
29	0,62	3,19	9,02	16,75	49,86	154,57	277,89	461,93	580,76	680,51	674,11
31	0,7	3,77	12,86	15,37	61,25	193,78	336,8	563,31	697,57	796,76	769,87
32	1,04	4,34	16,14	19,2	74,75	214,11	378,5	610,07	764,14	851,55	821,65
37	1,5	6,07	25,47	35,85	125,66	371,39	606,57	910,34	1093,55	1182,72	1099,34
41	3,35	9,35	36,85	53	183,49	516,49	806,03	1208,17	1386,07	1467,82	1348,77
43	3,88	11,28	39,59	59,07	232,78	579,51	929,99	1355,08	1540,01	1622,52	1485,22
47	4,81	19,13	61,99	89,73	300,46	786,42	1189,37	1700,65	1896,14	1954,43	1774,15
49	5,43	17,32	74,17	114,21	366,43	928,02	1358,69	1882,49	2090,92	2134,13	1928,06
53	6,3	27,09	101,06	153,09	455,3	1163,88	1693,28	2281,69	2485,59	2507,46	2256,42
59	11,51	41,96	156,79	216,85	678,71	1579,01	2267,93	2955,53	3140,44	3120,72	2790,71
61	12,04	44,76	182,73	261,8	754,29	1734,47	2482,16	3215,45	3361,17	3339,61	2984,66
64	17,55	59,59	210,21	294,46	900,45	1986,15	2848,01	3571,61	3726,2	3678,47	3283,98
67	20,46	71,03	243,34	359,28	1047,53	2287,66	3215,62	3987,72	4114,1	4037,8	3600,16
71	22,12	79,19	301,31	482,84	1296,54	2736,72	3735,67	4501,55	4634,38	4537,72	4041,74
73	25,8	97,01	316,92	497,39	1383,29	3014,34	4014,66	4796,66	4898,95	4797,19	4273,63
79	35,65	128,21	443,5	679,44	1803,98	3755,34	4951,83	5703,21	5766,61	5620,28	5003,15
81	38,3	145,78	515,62	716,65	1933,99	4048,4	5224,55	6009,57	6065,37	5912	5261,58

83	42,89	164,86	544,07	765,83	2107,8	4331,02	5614,15	6350,99	6378,22	6205,5	5520,34
89	61,25	210,48	713,01	1013,27	2573,4	5272,98	6602,92	7343,61	7340,23	7136,56	6348,54
97	75,2	285,94	968,95	1343,71	3471,17	6669,36	8129,83	8798,12	8737,72	8476,01	7542,36
101	91,89	312,4	1118,04	1587,54	3863,7	7459,78	8951,61	9596,87	9484,44	9187,88	8171,95
103	99,02	369,38	1189,33	1662,15	4107,96	7825,98	9364,76	9974,02	9860,97	9557,49	8499,81
107	132,61	445,43	1386,85	1884,15	4739,09	8640,6	10161,8	10784,4	10647,3	10314,7	9171,93
109	127,2	450,44	1512,04	2088,78	4928,47	9069,06	10668	11204,8	11051,5	10701,6	9517,37
113	138,11	533,45	1651,22	2280,89	5566	9989,25	11517,8	12059,2	11873,9	11502,9	10229,4
121	188,17	685,28	2118,54	2883,43	6921,08	11801,5	13496,1	13841,4	13621,8	13190,5	11727,7
125	217,34	801,51	2305,76	3200,62	7663,61	12797,9	14408,1	14779,2	14535,1	14073	12515,4
127	236,83	807,73	2430,54	3355,1	7941,66	13353	14989,1	15260,6	15004,3	14531,5	12918,3
128	245,31	810,73	2509,12	3360,07	8160,64	13545,3	15237	15530,2	15238,1	14755,5	13121,9
131	260,56	930,11	2781,51	3706,94	8731,52	14384,7	16004,5	16270,6	15967,9	15458,6	13744,2
137	316,49	1086,38	3215,85	4378,71	10065,2	16060,6	17578,4	17812,8	17467,7	16905,3	15035,2
139	327,88	1131,46	3427,43	4618,91	10530,6	16570,3	18176,6	18337,8	17974,6	17400,3	15475,6
149	424,35	1474,63	4183,32	5802,06	12930,9	19507,1	20988,6	21074	20660,9	19998,1	17780,2
151	486,24	1555,42	4422,01	6120,83	13332	20137,2	21617,8	21639,2	21214,1	20534,3	18259
157	524,54	1809,63	5156,17	6957,91	15083,2	22051	23466,6	23412,2	22933,6	22198,8	19744,4
163	578,89	1983,55	5776,29	7982,07	16692,5	24043,7	25365,6	25233	24722,8	23927,6	21276,6
167	667,92	2222,48	6239,54	8578,25	17780,6	25363	26691,1	26487,1	25948,2	25117,7	22334,2
169	739,24	2250,64	6566,7	8848,83	18425,9	25942,6	27337,9	27129	26577,1	25720,7	22872,7
173	821,91	2478,48	7002,11	9558,18	19741,8	27413,8	28702,4	28439,9	27849,8	26954,9	23972,3
179	864,38	2830,47	7911,84	10782,9	21609,2	29532,7	30788,8	30441,7	29811,2	28853,8	25659,1
181	884,36	2910,37	8246,83	11149,9	22474,1	30370,4	31519,7	31127	30483,7	29500,5	26235,7
191	1092,69	3521,71	9880,82	13222	25854,4	34134	35141,2	34670,1	33941,6	32847,6	29206,8
193	1173,58	3626,6	10282,1	13672,5	26556,3	34940	35937,4	35398	34654,8	33543,5	29828
197	1189,9	3846,26	11086,9	14765,6	28166,2	36507,8	37426,9	36885,6	36106,9	34947,9	31075,7
199	1247,74	3923,39	11456,9	15222,8	28764,4	37275,4	38226,4	37631	36847,6	35662,8	31703,5
211	1567,65	4981,46	13805,1	18307,4	33849,9	42304,6	43042,1	42310,4	41421,2	40087,4	35645,1
223	1971,23	5940	16533,8	21770,4	38744,6	47562,1	48146,9	47261,5	46268,5	44780,9	39816,4
227	2098,8	6422,74	17661,2	23063,3	40600,9	49440,3	49894,2	48971,7	47939	46397,8	41250,1
229	2174,32	6708,09	18230,7	23852,7	41777,1	50365	50767,2	49837,9	48789,1	47215,9	41986,7
233	2351,02	7011	19318	24935,5	43579	52166,2	52590,1	51592,6	50506,3	48880,8	43469,6
239	2503,51	7854,6	20752,8	27057,7	46284,7	54913,7	55345	54281,8	53140,6	51433,2	45722,4
241	2561,85	8012,73	21545,3	27749,4	47100,6	55984,3	56265,3	55195,4	54036,7	52295,3	46496,4
243	2768,48	8250,45	21978,8	28245	48403,3	56945,8	57215,9	56111,7	54936,6	53172,3	47265
251	3126,64	9331,99	24650,9	31548,3	52205,8	60923,1	61068,6	59868,9	58616,1	56724,3	50432,4
256	3183,66	9856,38	26063,6	33408,8	54789	63435,4	63517,9	62277,4	60976,5	59003,9	52468,7
257	3242,69	10168,3	26783,5	33830,4	55320	63976,9	64029,4	62765,2	61450,1	59469,2	52880,3
263	3530,71	11055,9	28556,5	36193,9	58695,3	67103,8	67056	65733,6	64353,9	62273,9	55383,8
269	3762,71	11717,9	30245,2	38434,1	61877,5	70314,2	70151,9	68763,9	67317,4	65152,9	57925,2
271	3951,28	12253,8	31345,1	39500,1	63297,1	71364,1	71216,6	69788,5	68320,5	66123,6	58788,2
277	4250,62	13197,1	33490,5	42485,3	66415,5	74586	74407,9	72908,2	71379,2	69086,8	61417,8
281	4442,57	14116,9	35103,7	44020,5	68392,8	76835,3	76576,7	75031,6	73455,3	71095,4	63210,1
283	4652,2	14365,1	35831	44857,2	69770,2	77933,1	77676,9	76105,4	74504,9	72110,7	64113,5
289	4876,44	15210,6	38145	47572,3	73471,8	81370,5	81020,9	79364,9	77698,3	75202,9	66851,1
293	5206,1	16014,5	40057,6	49783,1	76027,3	83695,2	83285,6	81578,3	79871,4	77288	68718,3
307	6172,16	19123,8	46196,6	57056,1	84652,6	91966,2	91427,5	89562,2	87671,9	84862,9	75450,7
311	6468,75	19957	47764,2	58669,7	87262,1	94431,9	93820,4	91904,1	89975	87079,7	77422,5
313	6526,72	20442,2	48921,1	60067,2	88598,1	95698,6	95038	93096,4	91139,3	88207,5	78418
317	6981,86	21368,6	50587,2	62476,8	91006,6	98149,5	97482,2	95488,4	93481,8	90471,9	80436,2

Tableau 1.1. $d = 2$

$l \setminus p$	0,001	0,002	0,004	0,005	0,01	0,02	0,03	0,05	0,07	0,1	0,2
2	0	0	0	0,02	0	0,02	0,07	0,04	0,12	0,25	0,75
3	0	0	0	0,06	0,17	0,24	0,63	1,87	2,59	5,3	14,18
4	0	0,06	0,15	0,1	0,47	2,8	3,74	12,5	20,49	31,23	46,79
5	0,06	0,27	0,34	0,58	3,33	8,97	23,19	45,78	71,38	88,8	99,81
7	0,2	0,68	4,44	3,94	24,53	87,72	165,28	255,51	293,39	304,77	276,79
8	0,4	1,95	7,52	14,59	72,6	186,74	322,78	437,34	465,45	459,59	412,12
9	0,68	4,83	22,38	36,3	143,81	390,58	545,01	658,05	672,28	658,17	585,97
11	4,85	21,05	81,27	153,2	491,58	987,83	1190,16	1259,37	1239,57	1201,05	1069,92
13	8,64	68,55	327,13	484,13	1256,78	1972,1	2082,86	2089,43	2046,29	1981,98	1762,83
16	73,56	319,93	1214,89	1660,52	3175,74	3954,97	3966,76	3896,22	3814,02	3692,66	3285,26
17	110,97	505,77	1700,65	2171,28	4046,69	4779,81	4767,58	4672,3	4574,95	4427,97	3937,69
19	235,06	1068,25	3256,02	4159,95	6352,5	6721,59	6658,78	6520,98	6383,86	6181,46	5498,59
23	997,71	3556,95	8447,21	9713,07	11859,2	11923	11807,6	11566	11324,3	10963,2	9747,23
25	1752,72	6098,53	12123,9	13606,8	15397,4	15318,4	15163,2	14853,5	14538,9	14072,4	12520,2
27	2985,64	9183,75	16558,5	17958,8	19418,2	19296,8	19101,2	18706,7	18317,6	17728	15765,4
29	4755,47	13452,2	21749,8	23114,6	24115,3	23910,8	23666,8	23183,2	22694,2	21966,5	19535,2
31	7200,7	18299,1	27622,9	28750,3	29499,4	29202,8	28906,5	28314,3	27721,8	26829,4	23858,2
32	9312,75	21627,3	31090,9	31966,5	32445,2	32122,8	31793,5	31145,4	30489,2	29509,1	26237,1
37	21606,6	40835,6	49936	50268,5	50154,5	49653,6	49147,2	48136	47128,3	45608,5	40551,1
41	36911,2	61558,8	68483,4	68564,8	68240,1	67553,3	66871,9	65495,3	64111,6	62051,2	55165,5
43	46973,9	73311,1	78972,8	79116,1	78719,8	77927,4	77138,2	75551,3	73965,9	71582,4	63643,2
47	72117,5	99693,5	103415	103312	102794	101764	100728	98653,2	96585	93469,9	83101,8

Tableau 1.2. $d = 3$

$l \setminus p$	0,001	0,002	0,004	0,005	0,01	0,02	0,03	0,05	0,07	0,1	0,2
2	0	0	0	0	0	0,1	0,37	0,71	0,78	2,16	6,78
3	0,14	0,13	0,3	0,38	1,74	8,19	11,51	29,82	46,09	58,48	64,24
4	0,17	0,81	5,32	4,92	21,11	91,41	139,5	216,16	225,47	228,98	206,09
5	2,13	5,45	30,89	55,91	169,24	441,13	548,59	589,38	580,95	564,01	502,69
7	47,86	219,86	877,64	1185,1	2123,82	2346,38	2327,56	2284,81	2236,06	2165,1	1926,78
8	133,53	801,91	2433,51	2888,01	3940,59	4012,48	3976,39	3896,22	3814,6	3692,51	3284,05
9	596,72	2569,43	5364,43	5838,67	6466,7	6432,47	6368,33	6238,3	6107,27	5912,26	5260,87
11	5178,3	11037,3	14397,6	14385,7	14498,1	14354,1	14206,7	13916,4	13625,4	13187,1	11729,8
13	18092,8	26979	28340,1	28423,3	28282,7	27997,6	27712,5	27144,4	26574,4	25721,7	22870,9
16	61074,2	65285,5	65279,2	65213,3	64888,5	64232,4	63584	62275	60967,9	59009,5	52460,6
17	80544	83309,4	83192,9	83110,7	82694,1	81860,8	81032,2	79365,7	77696,3	75201,5	66858,4
19	129578	130065	129806	129678	129030	127730	126434	123828	121231	117327	104299

Tableau 1.3. $d = 4$

$l \setminus p$	0,001	0,002	0,004	0,005	0,01	0,02	0,03	0,05	0,07	0,1	0,2
2	0	0	0,02	0,04	0,24	0,75	1,62	3,58	7,78	14,34	23,03
3	0,1	0,86	5,7	4,5	34,01	113,79	176,83	213,82	222,38	218,67	195,27
4	9,1	38,87	190,18	264,14	718,24	976,7	988,37	972,84	954,56	924,26	822,7
5	214,84	785,67	2206,48	2473,28	3063,24	3063,17	3034,24	2972,09	2911,34	2816,88	2507,72
7	10716,7	15988,1	16741,3	16726	16643,8	16476,5	16311	15975,5	15641,7	15139,2	13461,8
8	30527,1	32663,2	32639,5	32607,8	32446,7	32120,9	31791,7	31141,6	30485,5	29510,2	26237
9	58465,3	58933,9	58816,7	58759,2	58466,3	57881,3	57290,2	56116,6	54934,8	53168,8	47277,8

11	160894	160734	160419	160256	159453	157850	156252	153027	149806	144982	128897
----	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Tableau 1.4. $d = 5$

$l \setminus p$	0,001	0,002	0,004	0,005	0,01	0,02	0,03	0,05	0,07	0,1	0,2
2	0,14	0,2	0,75	0,24	2,15	5,03	12,95	24,34	35,79	50,39	50,77
3	11,39	22,53	107,51	171,38	519,44	688,45	700,32	692,02	678,15	658,26	585,64
4	574,71	2103,44	3806,65	3900,6	4055,17	4016,92	3978,17	3895,08	3815,53	3691,23	3284,9
5	12254,8	15343,4	15562,9	15550	15472,3	15319	15164,2	14852,7	14540,7	14072,7	12513,4
7	117535	117418	117187	117068	116483	115313	114143	111788	109441	105927	94166,1

Tableau 1.5. $d = 6$

CONCLUSION

Selon nous, la mise en application de l'estimateur de Monte Carlo que nous présentons dans le chapitre 4 devrait donner lieu à des applications intéressantes non seulement en génomique mais éventuellement aussi en biostatistique. Dans cette optique, il y aurait lieu selon nous de transporter cette mise en application vers la plate-forme SunSolaris. Concrètement, cela permettrait des tâches de fond, sans compter une puissance de calcul de beaucoup supérieure à celle d'un micro-ordinateur. À cette fin, il s'agirait donc d'adapter le texte source, le dialecte du langage C++ qu'on trouve sur micro-ordinateur étant sensiblement différent de celui qu'on trouve sur la plate-forme SunSolaris. On serait ainsi mieux à même d'effectuer des simulations.

Par ailleurs, nous pensons avoir trouvé une faille dans un des articles auxquels nous nous sommes intéressés ici. Il y aurait donc lieu selon nous de refaire l'analyse qu'on trouve dans cet article à la lumière de ce que nous présentons ici là-dessus.

Finalement, nous proposons de nouvelles espèces de structures, et, plus particulièrement, nous proposons, comme cas particulier du concept d'espèce évaluée, le concept d'espèce probabilisée. Dans cette optique, nous pensons qu'il serait intéressant de découvrir de nouvelles relations entre la statistique et la combinatoire et, plus particulièrement, entre la statistique et la théorie des espèces de structures. Plus précisément, il serait intéressant de découvrir de nouvelles relations entre ces deux disciplines-là qui passent par le biais de la théorie des espèces de structures.

BIBLIOGRAPHIE

- Autebert, J.-M. 1987. *Langages algébriques*. Paris : Masson.
- Balding, D. J., W. J. Bruno, E. Knill et D. C. Torney. 1996. *A comparative survey of non-adaptive pooling designs*. Coll. Speed, T. et M. S. Waterman, éditeurs, Coll. « *Genetic mapping and DNA sequencing* ». T. 81, série *The IMA Volumes in Mathematics and its Applications*, p. 133–154. Springer.
- Balding, D. J. et D. C. Torney. 1996. « Optimal pooling designs with error detection », *Journal of Combinatorial Theory, Série A*, vol. 74, p. 131–140.
- Barillot, E., D. Cohen et B. Lacroix. 1991. « Theoretical analysis of library screening using a n-dimensional pooling strategy », *Nucleic Acids research*, vol. 19, no. 22, p. 6241–6247.
- Berge, C. 1968. *Principes de combinatoire*. Paris : Dunod.
- . 1970. *Graphes et hypergraphes*. Paris : Dunod.
- Bergeron, F., G. Labelle et P. Leroux. 1994. *Théorie des espèces et combinatoire des structures arborescentes*. Montréal : Publications du LaCIM.
- Bernot, A. 1996. *L'analyse des génomes*. Nathan.
- Bourbaki, N. 1970. *Algèbre*. Paris : Diffusion C.C.L.S.
- Brown, S. M. 2003. *Essential of medical genomics*. John Wiley & Sons.
- Brown, T. A. 2004. *Génomes*. Coll. « Médecine-Sciences ». Flammarion.
- Bruno, W. J., E. Knill, D. J. Balding, D. C. Bruce, N. A. Doggett, W. W. Sawhill, R. L. Stallings, C. C. Whittaker et D. C. Torney. 1995. « Efficient pooling designs for library screening », *Genomics*, vol. 26, p. 21–30.
- Casella, G. et R. L. Berger. 2002. *Statistical Inference*. Pacific Grove : Duxbury.
- Dorfman, R. 1943. « The detection of defective members of large populations », *Annals of Mathematical Statistics*, vol. 14, no. 4, p. 436–440.
- Du, D.-Z. et F. K. Hwang. 2000. *Combinatorial group testing and its applications*. World Scientific, 2^e édition.
- Gabrini, P. J. 2005. *Structures de données avancées avec la STL*. Loze-Dion.

- Godement, R. 1973. *Cours d'algèbre*. Paris : Hermann.
- Labelle, J. 1981. *Théorie des graphes*. Mont-Royal : Modulo Éditeur.
- Luchetta, P., M.-C. Maurel, D. Higuët et M. Vervoort. 2005. *Évolution moléculaire*. Dunod.
- Olson, M., L. Hood, C. Cantor et D. Botstein. 1989. « A common language for physical mapping of the human genome », *Science*, vol. 245, p. 1434–1435.
- Park, H., W. Wu, Z. Liu, X. Wu et H. G. Zhao. 2003. « DNA screening, pooling design and simplicial complex », *Journal of Combinatorial Optimization, Série A*, vol. 7, p. 389–394.
- Ross, S. M. 1987. *Initiation aux probabilités*. Lausanne : Presses polytechniques romandes.
- Rotman, J. J. 1973. *The Theory of Groups*. Boston : Allyn and Bacon.
- Torney, D. C. 1999. « Sets pooling designs », *Annals of Combinatorics*, vol. 3, p. 95–101.